DELHI COLLEGE OF ENGINEERING

सत्यमेव जयते

LIBRARY

Class No. 510

Book No. W4l

Accession No. 45,234

Advanced Engineering Mathematics

C. R. WYLIE, JR.
Professor and Chairman, Department of Mathematics,
University of Utah

THIRD EDITION

INTERNATIONAL STUDENT EDITION

McGRAW-HILL BOOK COMPANY
New York
St. Louis
San Francisco
Toronto
London
Sydney
KŌGAKUSHA COMPANY, LTD.
Tokyo

Advanced Engineering Mathematics

INTERNATIONAL STUDENT EDITION

Exclusive rights by Kōgakusha Co., Ltd for manufacture and export from Japan This book cannot be re-exported from the country to which it is consigned by Kōgakusha Co., Ltd. or by McGraw-Hill Book Company or any of its subsidiaries

I

Copyright © 1960, 1966 by McGraw-Hill, Inc. All Rights Reserved. Copyright 1951 by McGraw-Hill, Inc. All Rights Reserved. This book, or parts thereof, may not be reproduced in any form without permission of the publishers.

Library of Congress Catalog Card Number 65-26769

TOSHO PRINTING CO., LTD., TOKYO, JAPAN

Preface

The first edition of this book was written with the announced purpose of providing an introduction to those branches of mathematics with which the average analytical engineer or physicist should be reasonably familiar in order to carry on his own work effectively and keep abreast of current developments in his field. In the present edition, as in the second, although the material has been completely rewritten, the objective remains the same, and the various additions, deletions, and refinements have been made only because they seemed to contribute to the realization of this goal.

Because ordinary differential equations are probably the most immediately useful part of postcalculus mathematics for the student of applied science and because the techniques of solving simple ordinary differential equations stem naturally from the techniques of calculus, the chapter on determinants and matrices with which the second edition began has been made a later chapter, and the book now begins with a chapter on ordinary differential equations of the first order This is followed by two other chapters on differential equations which develop the subject as far as the solution of systems of simultaneous linear equations with constant coefficients Following these is a chapter on finite differences containing not only the usual applications to interpolation, numerical differentiation and integration, and the step-by-step solution of differential equations, but also a section on linear difference equations with constant coefficients paralleling closely the preceding development for differential equations. This chapter also includes a discussion of curve fitting and the smoothing of data, as well as the method of least squares and the related topic of orthogonal polynomials One innovation in the present edition is the introduction of the Runge-Kutta method in addition to Milne's method for the step-by-step solution of differential equations. It is hoped that the material in this chapter will provide a useful background in classical finite differences on which a more extensive course in numerical analysis may be based. The fifth chapter is devoted to the application of the foregoing theory to mechanical and electrical systems, and, as in the first two editions, the mathematical identity of the two fields is emphasized However, a detailed discussion of the construction of electromechanical analogies is no longer included. The next two chapters deal, respectively, with Fourier series and integrals and with the Laplace transform, very much as did the corresponding chapters in the second edition. The chapters on separable partial differential equations and Bessel functions follow closely the development in the second edition, although many of the examples are new

The material on determinants and matrices which formed Chapter 1 in the second edition now appears substantially expanded as two chapters which follow the material on differential equations and related topics Next comes the chapter on vector analysis which, except for minor changes, is essentially the same as in the second edition Following the chapter on vector analysis is a new chapter devoted to an introduction to tensor analysis The last four chapters cover the theory of functions of a complex variable very much as did the corresponding chapters in the second edition

The book as presently organized falls naturally into three major subdivisions. The first nine chapters constitute a reasonably self-contained treatment of ordinary and partial differential equations and their applications The next four chapters cover the related areas of linear algebra, vector analysis, and tensor analysis; and the last four chapters cover the elementary theory and applications of functions of a complex variable With this organization, the book, which contains enough material for a two-year postcalculus course in applied mathematics, is well adapted to use as a text for any of several shorter courses

In the third edition, as in the first two, every effort has been made to keep the presentation detailed and clear while at the same time maintaining acceptable standards of precision and accuracy. To achieve this, more than the usual number of worked examples and carefully drawn figures have been included, and in every development there has been a conscious attempt to make the transitions from step to step so clear that a student with no more than a good background in calculus should seldom be held up more than momentarily. Over 400 new exercises of varying degrees of difficulty have been added to the problems already in the second edition. These range from formal problems of a purely routine nature to practical applications of considerable complexity. Hints are included in many of the exercises, and answers to the odd-numbered ones are given at the end of the book As in the first two editions, words and phrases defined in the body of the text are set in boldface, and italic type has been liberally used as a sign of emphasis. Theorems, corollaries, and formal definitions are set on wider lines than the main body of the text, and illustrative examples are set in type of a different size

The indebtedness of the author to his colleagues, students, and former teachers is too great to catalog, and to all who have given help and encouragement in the preparation of this book, I can offer here only a most inadequate acknowledgment of my appreciation In particular, I am deeply grateful to those users of this book who have been kind enough to write me their impressions and criticisms of the first two editions and their suggestions for an improved third edition. Finally, I must express my gratitude to my wife, Ellen, and to my secretary, Mrs Jason Everts, who gave me invaluable assistance in proofreading the manuscript.

C. R. WYLIE, JR.

To the Student

This book has been written to help you in your development as an applied scientist, whether engineer, physicist, chemist, or mathematician. It contains material which you will find of great use, not only in the technical courses you have yet to take, but also in your profession after graduation as long as you deal with the analytical aspects of your field.

I have tried to write a book which you will find not only useful but also easy to study from, at least as easy as a book on advanced mathematics can be. There is a good deal of theory in it, for it is the theoretical portion of a subject which is the basis for the nonroutine applications of tomorrow. But nowhere will you find theory for its own sake, interesting and legitimate as this may be to a pure mathematician. Our theoretical discussions are designed to illuminate principles, to indicate generalizations, to establish limits within which a given technique may or may not safely be used, or to point out pitfalls into which one might otherwise stumble. On the other hand, there are many applications illustrating, with the material at hand, the usual steps in the solution of a physical problem: formulation, manipulation, and interpretation. These examples are, without exception, carefully set up and completely worked, with all but the simplest steps included. Study them carefully, with paper and pencil at hand, for they are an integral part of the text. If you do this you should find the exercises, though challenging, still within your ability to work.

There are two minor points of notation which, when appreciated, should add to the ease with which you can read this book. In the first place, concepts and terms defined in formal definitions and terms defined informally in the body of the text are always indicated by the use of **boldface type.** Second, *italic type* is used in various places as a sign of emphasis, corresponding to the stress which a person would place upon key words when speaking. One final suggestion to you in your study of this book is that you read each section through for the main ideas before you concentrate on filling in any of the details You will probably be surprised at how many times a detail which seems to hold you up in one paragraph is explained in the next as the discussion unfolds

Because this book is a long one and contains material suitable for various courses, your teacher may begin with any of a number of chapters. However, the overall structure of the book is the following: The first nine chapters are devoted to the general theme of ordinary and partial differential equations and related topics Here you will find basic analytical techniques for solving the equations in which physical problems must be formulated when continuously changing quantities are involved. Chapters 10 through 13 deal with the somewhat related topics of linear algebra, vector analysis, and an introduction to generalized coordinates and tensor analysis. Finally, Chapters 14 through 17 provide an introduction to the theory and applications of functions of a complex variable (Chapter 4, in particular, is worthy of note because it provides an introduction to numerical analysis, the modern field which deals with techniques for obtaining numerical answers to problems too complicated to be solved by exact analytic methods)

It has been gratifying to me to receive from time to time letters from students who have used this book, giving me their reactions to it, pointing out errors and misprints in it, and offering suggestions for its improvement Should you be inclined to do so, I should be happy to hear from you also And now good luck and every success.

C R. WYLIE, JR

Contents

Preface v
To the Student ix

chapter 1

Ordinary differential equations of the first order 1

1.1 Introduction 1
1.2 Fundamental Definitions 2
1.3 Separable First-order Equations 8
1.4 Homogeneous First-order Equations 11
1.5 Exact First-order Equations 14
1.6 Linear First-order Equations 19
1.7 Applications of First-order Differential Equations 21

chapter 2

Linear differential equations with constant coefficients 30

2.1 The General Linear Second-order Equation 30
2.2 The Homogeneous Linear Equation with Constant Coefficients 36
2.3 The Nonhomogeneous Equation 42
2.4 Particular Integrals by the Method of Variation of Parameters 49
2.5 Equations of Higher Order 52
2.6 Applications 55

chapter 3

Simultaneous linear differential equations 66

3.1 Introduction 66
3.2 The Reduction of a System to a Single Equation 67
3.3 Complementary Functions and Particular Integrals for Systems of Equations 74

chapter 4
Finite differences 79
4.1 The Differences of a Function 79
4.2 Interpolation Formulas 90
4.3 Numerical Integration and Differentiation 99
4.4 The Numerical Solution of Differential Equations 108
4.5 Difference Equations 117
4.6 The Method of Least Squares 126

chapter 5
Mechanical and electrical circuits 144
5.1 Introduction 144
5.2 Systems with One Degree of Freedom 144
5.3 The Translational-mechanical System 151
5.4 The Series-electrical Circuit 165
5.5 Systems with Several Degrees of Freedom 171

chapter 6
Fourier series and integrals 181
6.1 Introduction 181
6.2 The Euler Coefficients 182
6.3 Half-range Expansions 189
6.4 Alternative Forms of Fourier Series 196
6.5 Applications 200
6.6 Harmonic Analysis 206
6.7 The Fourier Integral as the Limit of a Fourier Series 211
6.8 From the Fourier Integral to the Laplace Transform 222

chapter 7
The Laplace transformation 226
7.1 Theoretical Preliminaries 226
7.2 The General Method 232
7.3 The Transforms of Special Functions 237
7.4 Further General Theorems 242
7.5 The Heaviside Expansion Theorems 255
7.6 Transforms of Periodic Functions 260
7.7 Convolution and the Duhamel Formulas 270

chapter 8
Partial differential equations 282
8.1 Introduction 282
8.2 The Derivation of Equations 282
8.3 The D'Alembert Solution of the Wave Equation 294
8.4 Separation of Variables 302
8.5 Orthogonal Functions and the General Expansion Problem 311
8.6 Further Applications 328
8.7 Laplace Transform Methods 338

chapter 9

Bessel functions and Legendre polynomials 345

9.1 Theoretical Preliminaries 345
9.2 The Series Solution of Bessel's Equation 351
9.3 Modified Bessel Functions 357
9.4 Equations Reducible to Bessel's Equation 363
9.5 Identities for the Bessel Functions 365
9.6 Orthogonality of the Bessel Functions 372
9.7 Applications of Bessel Functions 377
9.8 Legendre Polynomials 388

chapter 10

Determinants and matrices 400

10.1 Determinants 400
10.2 Elementary Properties of Matrices 415
10.3 Adjoints and Inverses 429
10.4 Rank and the Equivalence of Matrices 437
10.5 Systems of Linear Equations 444
10.6 Matric Differential Equations 461

chapter 11

Further properties of matrices 466

11.1 Quadratic Forms 466
11.2 The Characteristic Equation of a Matrix 477
11.3 The Transformation of Matrices 492
11.4 Functions of a Square Matrix 505
11.5 The Cayley-Hamilton Theorem 517
11.6 Infinite Series of Matrices 525

chapter 12

Vector analysis 532

12.1 The Algebra of Vectors 532
12.2 Vector Functions of One Variable 545
12.3 The Operator ∇ 550
12.4 Line, Surface, and Volume Integrals 559
12.5 Integral Theorems 572
12.6 Further Applications 585

chapter 13

Tensor analysis 595

13.1 Introduction 595
13.2 Oblique Coordinates 595
13.3 Generalized Coordinates 605
13.4 Tensors 619
13.5 Divergence and Curl 624
13.6 Covariant Differentiation 628

chapter 14

Analytic functions of a complex variable 633

14.1 Introduction 633
14.2 Algebraic Preliminaries 633
14.3 The Geometric Representation of Complex Numbers 636
14.4 Absolute Values 641
14.5 Functions of a Complex Variable 644
14.6 Analytic Functions 650
14.7 The Elementary Functions of z 656
14.8 Integration in the Complex Plane 663

chapter 15

Infinite series in the complex plane 676

15.1 Series of Complex Terms 676
15.2 Taylor's Expansion 686
15.3 Laurent's Expansion 692

chapter 16

The theory of residues 699

16.1 The Residue Theorem 699
16.2 The Evaluation of Real Definite Integrals 704
16.3 The Complex Inversion Integral 711
16.4 Stability Criteria 716

chapter 17

Conformal mapping 729

17.1 The Geometrical Representation of Functions of z 729
17.2 Conformal Mapping 732
17.3 The Bilinear Transformation 737
17.4 The Schwarz-Christoffel Transformation 748

Appendix 755

Graeffe's Root-squaring Process 755

Answers to odd-numbered exercises 765

Index 801

CHAPTER ONE

Ordinary Differential Equations of the First Order

1.1

Introduction

An equation involving one or more derivatives of a function is called a **differential equation.** By a **solution** of a differential equation is meant a relation between the dependent and independent variables which is free of derivatives and which, when substituted into the given equation, reduces it to an identity The study of the existence, nature, and determination of solutions of differential equations is of fundamental importance not only to the pure mathematician but also to anyone engaged in the mathematical analysis of natural phenomena

In general, a mathematician considers it a triumph if he is able to prove that a given differential equation possesses a solution and if he can deduce a few of the more important properties of that solution A physicist or engineer, on the other hand, is usually greatly disappointed if a specific expression for the solution cannot be exhibited The usual compromise is to find some practical procedure by means of which the required solution can be approximated with satisfactory accuracy.

Not all differential equations are of such difficulty as to make this necessary, however, and there are several large and very important classes of equations for which solutions can readily be found For instance, an equation such as

$$\frac{dy}{dx} = f(x)$$

is really a differential equation, and the integral

$$y = \int f(x)\,dx + c$$

is a solution More generally, the equation

$$\frac{d^n y}{dx^n} = g(x)$$

is a differential equation whose solution can be found by n successive integrations. Except in name, the process of integration is actually an example of a process for solving differential equations.

In this and the following two chapters we shall consider those differential equations which are next in difficulty after those which can be solved by direct integration These equations form only a very small part of the class of all differential equations, and yet with a knowledge of them a scientist is equipped to handle a great variety of applications To get so much for so little is indeed remarkable!

1.2

Fundamental definitions

If the derivatives which appear in a differential equation are total derivatives, the equation is called an **ordinary differential equation**; if partial derivatives occur, the equation is called a **partial differential equation.** By the **order** of a differential equation is meant the order of the highest derivative which appears in the equation

EXAMPLE 1

The equation $x^2y'' + xy' + (x^2 - 4)y = 0$ is an *ordinary* differential equation of the *second* order connecting the dependent variable y with its first and second derivatives and with the independent variable x

EXAMPLE 2

The equation $\dfrac{\partial^4 u}{\partial x^4} + 2\dfrac{\partial^4 u}{\partial x^2\,\partial y^2} + \dfrac{\partial^4 u}{\partial y^4} = 0$ is a *partial* differential equation of the *fourth* order

At present we shall be concerned exclusively with ordinary differential equations

An equation which is linear, that is, of the first degree, in the *dependent* variable and its derivatives is called a **linear differential equation.** All other equations are called **nonlinear.** In general, linear equations are much easier to solve than nonlinear ones, and most elementary applications involve linear equations

EXAMPLE 3

The equation $y'' + 4xy' + 2y = \cos x$ is a *linear* equation of the *second* order The presence of the terms xy' and $\cos x$ does not alter the fact that the equation is linear, because, by definition, linearity is determined solely by the way the *dependent* variable y and its derivatives enter into combination among themselves

EXAMPLE 4

The equation $y'' + 4yy' + 2y = \cos x$ is a *nonlinear* equation because of the occurrence of the product of y and one of its derivatives

EXAMPLE 5

The equation $y'' + \sin y = 0$ is *nonlinear* because of the presence of $\sin y$, which is a nonlinear function of y

As illustrated by the simple equation

$$\frac{dy}{dx} = e^{-x^2}$$

and its solution

$$y = \int e^{-x^2}\,dx + c$$

the solution of a differential equation may depend upon integrals which cannot be evaluated in terms of elementary functions. This example also illustrates the fact that a solution of a differential equation usually involves one or more arbitrary constants.

A detailed treatment of the question of the maximum number of *essential* arbitrary constants that a general solution of a differential equation may contain or even of what is meant by essential constants is quite difficult * For our purposes, if an expression contains n arbitrary constants we shall consider them essential if they cannot, through formal rearrangement of the expression, be replaced by any smaller number of constants For example,

$$a\cos^2 x + b\sin^2 x + c\cos 2x \tag{1}$$

contains three arbitrary constants However, since

$$\cos 2x = \cos^2 x - \sin^2 x$$

the expression (1) can be written in the form

$$\begin{aligned} a\cos^2 x + b\sin^2 x + c(\cos^2 x - \sin^2 x) &= (a + c)\cos^2 x + (b - c)\sin^2 x \\ &= d\cos^2 x + e\sin^2 x \end{aligned}$$

where $d = a + c$ and $e = b - c$ The fact that the three arbitrary constants a, b, and c can be replaced by the two constants d and e shows that the former are not all essential. On the other hand, since $\cos^2 x$ and $\sin^2 x$ are linearly independent† (whereas $\cos^2 x$, $\sin^2 x$, and $\cos 2x$ are linearly dependent), it follows that there is no further rearrangement of the given expression that will permit d and e to be combined into and replaced by a single new arbitrary constant Hence d and e are essential

It is frequently the case (especially with linear equations) that a differential equation of order n possesses solutions containing n essential arbitrary constants but none containing more.

---

* See, for instance, R. P Agnew, "Differential Equations," 2d ed , pp 103–105, McGraw-Hill Book Company, New York, 1960

† See Definitions 1 and 2, Sec 10 5

However, there are equations such as

$$\left|\frac{dy}{dx}\right| + |y| = 0$$

(which has only the single solution $y = 0$) and

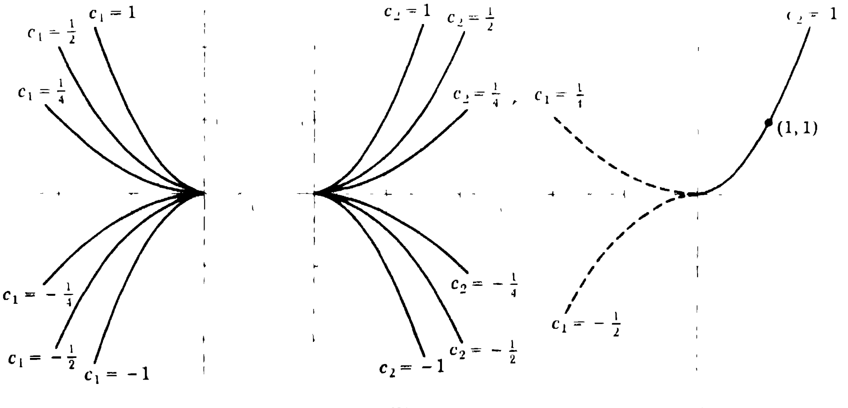
$$\left|\frac{dy}{dx}\right| + 1 = 0$$

(which has no solutions at all) which possess *no* solutions containing *any* arbitrary constants. Moreover, there are also simple differential equations which possess solutions containing more essential parameters than the order of the equation. For instance, it is easy to verify that the arc of the family $y = c_1x^2$ $(x \leqq 0)$ (Fig. 1.1*a*) corresponding to any value of c_1 can be paired with the arc of the family $y = c_2x^2$ $(x \geqq 0)$ (Fig. 1.1*b*) corresponding to any value of c_2, to give a function which satisfies the differential equation

$$xy' = 2y \tag{2}$$

for all values of x (Fig. 1.1*c*). A still more striking example of this sort appears in Exercise 30, where a first-order equation with a solution containing infinitely many essential parameters is given.

As the foregoing suggests, it is difficult, if not impossible, to make statements valid for all differential equations. The theory of differential equations is essentially a body of theorems concerning particular classes of equations defined by such considerations as linearity, order, and continuity. Typical of these is the follow-



(*a*) (*b*) (*c*)

FIGURE 1.1

Arcs of different parabolas of the family $y = cx^2$ pieced together to give solutions of the differential equation $xy' = 2y$

ing result,[*] which is of fundamental importance in the study of the equations we shall consider in this chapter, namely, equations of the first order

THEOREM 1

Let (x_0,y_0) be a point of the xy-plane, let R be the rectangular region defined by the inequalities $|x - x_0| \leqq a$, $|y - y_0| \leqq b$; let $f(x,y)$ and $f_y(x,y) = \dfrac{\partial f(x,y)}{\partial y}$ be single-valued and continuous at all points of R, let M be a constant such that $|f(x,y)| < M$ at all points of R, and let h be the smaller of the numbers a and b/M Then, on the interval $|x - x_0| < h$, there is a unique continuous function y which satisfies the equation $y' = f(x,y)$ and which takes on the value y_0 when $x = x_0$

It is instructive to reconsider Eq (2) in the light of Theorem 1 For this equation we have $f(x,y) = 2y/x$, and, clearly, neither f nor f_y exists when $x = 0$ Hence, it follows from Theorem 1 that, over an interval containing $x = 0$, neither the existence nor the uniqueness of a solution of Eq (2) can be guaranteed Actually, as our earlier discussion pointed out, Eq (2) does have solutions which are valid for all values of x However, as Fig 1 1c illustrates, over any interval which contains $x = 0$, the solution curve which passes through a given point (x_0,y_0), e.g , (1,1), is not unique On the other hand, according to Theorem 1, over any interval which contains x_0 but does not contain $x = 0$, the solution curve which passes through a given point (x_0,y_0) is unique

Almost all applications of differential equations involve equations which possess solutions containing at least one arbitrary constant, and for such equations it is convenient to introduce the following definitions A solution which contains at least one arbitrary constant is called a **general solution.** A solution obtained from a general solution by assigning particular values to the arbitrary constants which appear in it is called a **particular solution.** Solutions which cannot be obtained from any general solution by assigning specific values to its arbitrary constants are called **singular solutions.** If a general solution has the property that every solution of the differential equation can be obtained from it by assigning suitable values to its arbitrary constants, it is said to be a **complete solution.** A general solution can thus be thought of as a description of some family of particular solutions, and a complete solution can be thought of as a description of the set of all solutions of the given equation

It is important to note that we speak of *a* general solution and *a* complete solution of a differential equation and not of *the* general solution and *the* complete solution. If an equation has a general solution or a complete solution, it has many such solutions, and these may differ significantly in form. Moreover, in particular problems involving differential equations, the choice of which complete solution to use often has an important bearing on the ease with which the problem can be solved.

---

[*] See, for instance, M Golomb and M E Shanks, "Elements of Ordinary Differential Equations," 2d ed., pp 63–78, McGraw-Hill Book Company, New York, 1965

EXAMPLE 6

Verify that $y = ae^{-x} + be^{2x}$ is a solution of the equation $y'' - y' - 2y = 0$ for all values of the constants a and b.

By differentiating y, substituting into the differential equation as indicated, and then collecting terms on a and b, we obtain

$$\begin{aligned}(ae^{-x} + 4be^{2x}) - (-ae^{-x} + 2be^{2x}) - 2(ae^{-x} + be^{2x})&\\ = (e^{-x} + e^{-x} - 2e^{-x})a + (4e^{2x} - 2e^{2x} - 2e^{2x})b&\\ = 0 \cdot a + 0 \cdot b = 0&\end{aligned}$$

for all values of a and b. Thus, $y = ae^{-x} + be^{2x}$ is a general solution of $y'' - y' - 2y = 0$. In fact, as we shall see in Sec. 2.2, it is a complete solution of this equation.

It is interesting to note that, although $y_1 = ae^{-x}$ and $y_2 = be^{2x}$ also satisfy the equation $yy'' - (y')^2 = 0$ for all values of a and b, the sum

$$y = y_1 + y_2 = ae^{-x} + be^{2x}$$

is not a solution of $yy'' - (y')^2 = 0$. *In fact, differentiating, substituting, and simplifying, we have*

$$(ae^{-x} + be^{2x})(ae^{-x} + 4be^{2x}) - (-ae^{-x} + 2be^{2x})^2 = 9abe^{x}$$

and this cannot vanish identically unless either a or b is zero, that is, unless the sum y consists of just one or the other of the two individual solutions. Roughly speaking, the reason for this difference in behavior is that the equation $y'' - y' - 2y = 0$ is linear, whereas the equation $yy'' - (y')^2 = 0$ is nonlinear. More precisely, as we shall see in Theorem 1, Sec. 2.1, for linear equations in which y or one of its derivatives appears in every term, the sum of two solutions is also a solution, whereas, in general, the sum of two solutions of a nonlinear equation is not a solution.

Occasionally it is necessary to determine a differential equation of order n which has a given function containing n arbitrary constants as a general solution. This can be done (at least theoretically) by differentiating the given expression n times and then eliminating the arbitrary constants by algebraic manipulation of the resulting equations.

EXAMPLE 7

If a and b are arbitary constants, find a second-order equation which has

$$y = ae^{x} + b\cos x \tag{3}$$

as a general solution.

By differentiating the given expression, we find

$$y' = ae^{x} - b\sin x \tag{4}$$

$$y'' = ae^{x} - b\cos x \tag{5}$$

Then, by adding and subtracting Eqs. (3) and (5), we obtain

$$a = \frac{y + y''}{2e^{x}} \qquad b = \frac{y - y''}{2\cos x}$$

Substitution of these into Eq (4) gives

$$y' = \frac{y + y''}{2e^x} e^x - \frac{y - y''}{2 \cos x} \sin x$$

and finally

$$(6) \qquad (1 + \tan x)y'' - 2y' + (1 - \tan x)y = 0$$

Although Eq (6), except for its obvious multiples, is the only second-order differential equation having (3) as a general solution, it is by no means the only equation of which (3) is a general solution For instance, if (5) is differentiated twice more we obtain

$$y^{IV} = ae^x + b \cos x$$

and by comparing this with (3) we can see that the given function also satisfies the very simple equation

$$(7) \qquad y^{IV} = y$$

Since Eq. (7) is of the fourth order, it presumably possesses general solutions containing four arbitrary constants, and it is easy to verify that

$$y = ae^x + b \cos x + ce^{-x} + d \sin x$$

does in fact satisfy Eq (7)

EXERCISES

Describe each of the following equations, giving its order and telling whether it is ordinary or partial and linear or nonlinear

1 $y'' + 3(y')^2 + 4y = 0$

2 $y'' + (a + b \cos 2x)y = 0$

3 $y'' + y' + \cos y = 0$

4 $y''' + 6y'' + 4y' + y = e^x$

5 $\dfrac{d(xy')}{dx} + x^2y = 0$

6 $u \dfrac{\partial^2 u}{dx^2} = \dfrac{\partial^2 u}{\partial x \, \partial t}$

7 $\dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} + \dfrac{\partial^2 u}{\partial z^2} = 0$

8 $\dfrac{\partial^2 \left(x^2 \dfrac{\partial^2 y}{\partial x^2} \right)}{\partial x^2} = \dfrac{\partial^2 y}{\partial t^2}$

Verify that each of the following equations has the indicated solution for all values of a and b

9 $y'' - 6y' + 9y = 0$ $\qquad y = ae^{3x} + bxe^{3x}$

10 $y'' + 4y = 0$ $\qquad y = a \cos 2x + b \sin 2x$

11 $(\cos 2x)y' + (2 \sin 2x)y = 2$ $\qquad y = a \cos 2x + \sin 2x$

12 $y'' + 2y' + 2y = 0$ $\qquad y = e^{-x}(a \cos x + b \sin x)$

13 $2xy \, dy = (y^2 - x) \, dx$ $\qquad y^2 = ax - x \ln |x|$

14 $(xy - x^2) \, dy = y^2 \, dx$ $\qquad y = ae^{y/x}$

15 $y'' + (y')^2 + 1 = 0$ $\qquad y = \ln |\cos (x - a)| + b$

16 $\dfrac{\partial^2 u}{\partial x^2} = \dfrac{\partial u}{\partial t}$ $\qquad u = ae^{-9t} \cos (3x + b)$

17 $4\dfrac{\partial^2 u}{\partial x^2} = \dfrac{\partial^2 u}{\partial t^2}$ $\qquad u = af(x + 2t) + bg(x - 2t)$

If a and b are arbitrary constants, find a differential equation of minimum order of which each of the following expressions is a general solution

18 $y = ae^{-t} + be^t$

19 $y = ae^{-t} + be^t + ce^{2t}$

20 $y = ae^{-2t} + bte^{-2t}$

21 $y = 2ax + bx^2$

22 $y = e^{-x} + be^{2x}$

23 $y = a \cosh 2x + b \sinh 2x$

24 $y = \sin (ax + b)$

25 Find a differential equation which has as a general solution the expression which defines the family of all parabolas which touch the x-axis and have their axes vertical

26 Find a differential equation which has as a general solution the expression which defines the family of all lines which touch the parabola $2y = x^2$

27 Verify that, for all values of the arbitrary constants a and b, both $y_1 = ax^2$ and $y_2 = b(x - 1)^2$ satisfy each of the differential equations

$$(x^2 - x)y'' - (2x - 1)y' + 2y = 0 \qquad \text{and} \qquad 2yy'' = (y')^2$$

but that $y = ax^2 + b(x - 1)^2$ will satisfy only the first of these equations Explain

28 Verify that, for all values of the arbitrary constants a, b, m, n, both $y_1 = ae^{mx}$ and $y_2 = be^{nx}$ satisfy the nonlinear differential equation $y'''y - y''y' = 0$ Under what conditions, if any, does the sum $y = y_1 + y_2 = ae^{mx} + be^{nx}$ also satisfy this equation?

29 Verify that, for all values of the arbitrary constants c_1 and c_2, the differential equation $xy' - 2y + 2 = 0$ is satisfied by the function

$$y = \begin{cases} c_1x^2 + 1 & x \leqq 0 \\ c_2x^2 + 1 & x > 0 \end{cases}$$

Explain

30 Verify that, for all values of the arbitrary constants $\{c_n\}$ $(n = \quad , -2, -1, 0, 1, 2, \quad)$, the differential equation

$$(1 - \cos x)y' - (\sin x)y = 0$$

is satisfied by the function

$$y = c_n(1 - \cos x) \qquad 2n\pi \leqq x < 2(n + 1)\pi$$

Explain

1.3

Separable first-order equations

In many cases a first-order differential equation can be reduced by algebraic manipulations to the form

$$f(x)\,dx = g(y)\,dy \tag{1}$$

Such an equation is said to be **separable**, because the variables x and y can be *separated* from each other in such a way that x appears only in the coefficient of dx and y appears only in the coefficient of dy An equation of this type can be solved at once by integration, and we have the general solution

$$\int f(x)\,dx = \int g(x)\,dx + c \tag{2}$$

where c is an arbitrary constant of integration It must be borne in mind, however, that the integrals which appear in (2) may be impossible to evaluate in terms of elementary functions, and numerical or graphical integration may be required before this solution can be put to practical use.

Other forms which should be recognized as being separable are

$$f(x)G(y)\,dx = F(x)g(y)\,dy \tag{3}$$

$$\frac{dy}{dx} = M(x)N(y) \tag{4}$$

The general solution of Eq (3) can be found by first dividing by the product $F(x)G(y)$ to separate the variables and then integrating:

$$\int \frac{f(x)}{F(x)}\,dx = \int \frac{g(y)}{G(y)}\,dy + c$$

Similarly, a general solution of Eq. (4) can be found by first multiplying by dx and dividing by $N(y)$ and then integrating:

$$\int \frac{dy}{N(y)} = \int M(x)\,dx + c$$

Clearly, the process of solving a separable equation will often involve division by one or more expressions In such cases the results are valid where the divisors are not equal to zero, but may or may not be meaningful for values of the variables for which the division is impossible Such values require special consideration, and, as we shall see in the next example, may lead us to singular solutions

EXAMPLE 1

Solve the differential equation $dx + xy\,dy = y^2\,dx + y\,dy$

It is not immediately evident that this equation is separable In any case, however, the best first step in solving an equation of this sort is to collect terms on dx and dy This gives

$$(1 - y^2)\,dx = y(1 - x)\,dy$$

which is of the form (3) Hence, division by the product $(1 - x)(1 - y^2)$ will separate the variables and reduce the equation to the standard form (1)

$$\frac{dx}{1 - x} = \frac{y\,dy}{1 - y^2}$$

Now, multiplying by -2 and integrating, we obtain the following equation defining y as an implicit function of x

$$2 \ln |1 - x| = \ln |1 - y^2| + c$$

In this case, as in many problems of this sort, it is possible to write the solution in a more convenient form by first combining the logarithmic terms and then taking antilogs

$$\ln \frac{|1 - x|^2}{|1 - y^2|} = c \qquad \frac{|1 - x|^2}{|1 - y^2|} = e^c = k^2$$

where $k^2 = e^c$ is necessarily positive Finally, clearing of fractions and eliminating the absolute values, we have

$$(1 - x)^2 = \pm k^2(1 - y^2) \qquad k \neq 0$$

The two possibilities here can, of course, be combined into one by writing

$$(1 - x)^2 = \lambda(1 - y^2)$$

where now λ can take on any real value, positive or negative, except 0 The solution of the differential equation thus defines the family of conics

$$\frac{(x - 1)^2}{\lambda} + y^2 = 1 \qquad \lambda \neq 0 \tag{5}$$

typical members of which are shown in Fig 1 2 If $\lambda > 0$, the solution curves are all ellipses, if $\lambda < 0$, the solution curves are all hyperbolas

In most practical problems a general solution of a differential equation is required to satisfy specific conditions which permit its arbitrary constants to be uniquely determined. For instance, in the present problem we might ask for the particular solution curve which passes through the point $(-\frac{7}{5}, \frac{13}{5})$. Substituting these values of x and y, we then have

$$\frac{(-\frac{7}{5} - 1)^2}{\lambda} + \left(\frac{13}{5}\right)^2 = 1 \qquad \lambda = -1$$

and the specific solution

$$y^2 = 1 + (x - 1)^2 \tag{6}$$

Equation (6) defines that unique member of the family of curves (5) which passes through the point $(-\frac{7}{5}, \frac{13}{5})$. However, over any interval which contains $x = 1$, there are many functions which satisfy the given differential equation and are such that $y = \frac{13}{5}$ when $x = -\frac{7}{5}$. In fact, the upper branch of *any* curve of the family (5) for $x > 1$ can be associated with the upper branch of the curve (6) for $x \leqq 1$ to give a function which satisfies the given equation and fulfills the condition that $y = \frac{13}{5}$ when $x = -\frac{7}{5}$. This is, of course, consistent with the fact that, according to Theorem 1, Sec. 1.2, the uniqueness of the solution for which $y = \frac{13}{5}$ when $x = -\frac{7}{5}$ can be guaranteed only over an interval around $x = -\frac{7}{5}$ which does not contain $x = 1$, since y' is undefined at $x = 1$.

It should be noted that, in separating variables in the given equation, it was necessary to divide by $1 - x$ and by $1 - y^2$, hence, the possibility that $x = 1$ and the possibility that $y = \pm 1$ were implicitly ruled out. Therefore, had we desired the particular solution curve which passed through any point with coordinates of the form $(1,y_0)$, $(x_0,1)$, or $(x_0,-1)$, we could not have found that curve, if it existed at all, by using the general solution and particularizing λ. It would have been necessary, instead, to return to the differential equation and search for the required solution by some method other than separation of variables. In this case it is obvious that the

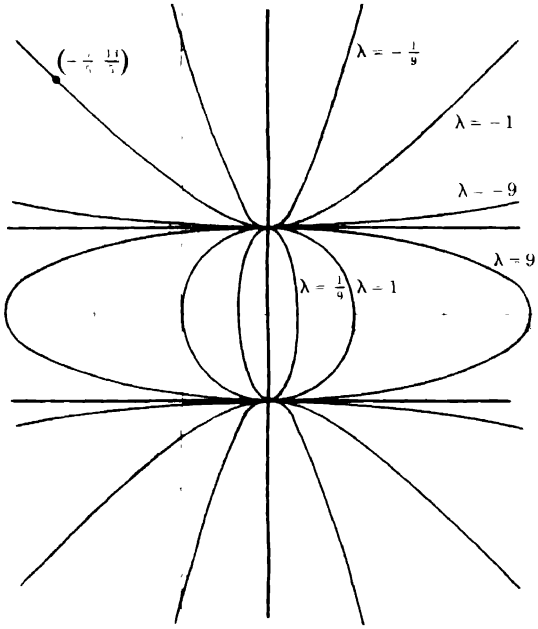


FIGURE 1.2 Typical members of the solution family $[(x - 1)^2/\lambda] + y^2 = 1$ of the differential equation $(1 - y^2)\,dx = y(1 - x)\,dy$

linear equations $x = 1$, $y = 1$, and $y = -1$ all define solutions of the given differential equation and, moreover, satisfy, respectively, the conditions $(1,y_0)$, $(x_0,1)$, and $(x_0,-1)$. None of these can be *obtained* from our general solution, although $x = 1$ can be *included* in the first form of it by permitting λ to take on the (previously excluded) value zero Hence $y = 1$ and $y = -1$ appear as singular solutions of the given equation

EXERCISES

Find a general solution of each of the following equations

1	$x\,dy = 3y\,dx$	2	$3x^2(1 + y^2)\,dx = dy$
3	$y\,dy = 2(xy + x)\,dx$	4	$y\,dx = 2(xy + x)\,dy$
5	$x\,dy = (y^2 - 3y + 2)\,dx$	6	$dx + y\,dy = x^2y\,dy$
7	$y^2\,dx - xy\,dy = xy(dy - y\,dx)$	8	$(xy^2 - x)\,dx = (y + x^2y)\,dy$
9	$ye^{x+y}\,dy = dx$	10	$yy'' = (y')^2$

Find that particular solution of each of the following equations which satisfies the indicated conditions:

11 $dy = x(2y\,dx - x\,dy)$ $\quad x = 1,\ y = 4$

12 $2x\,dx - dy = x(x\,dy - 2y\,dx)$ $\quad x = -3,\ y = 1$

13 Is there a solution of the equation $x\,dy = 3(y - 1)\,dx$ with the property that $y = 3$ when $x = 1$ and $y = 9$ when $x = 2$? Is there a solution of this equation with the property that $y = 3$ when $x = -1$ and $y = 9$ when $x = 2$? Explain

14 Find a solution of the equation $(1 - x^2)\,dy = -4xy\,dx$ with the property that $y = 9$ when $x = -2$, $y = 2$ when $x = 0$, and $y = 0$ when $x = 2$

15 Show that every solution of the equation $y' = ky$ is of the form $y = Ae^{kx}$ (Hint Let y be any solution of the given equation, and consider the derivative of the fraction y/e^{kx})

16 A critical student watching his professor integrate the separable equation $f(x)\,dx = g(y)\,dy$ objected that the procedure was incorrect, since one side was integrated with respect to x while the other side was integrated with respect to y How would you answer the student's objection?

17 Show that the change of dependent variable defined by the substitution $v = ax + by + c$ will always transform the equation $y' = f(ax + by + c)$ into a separable equation

Find a general solution of each of the following equations

18	$y' = (x - y)^2$	19	$y' = -2 + e^{2x+y-1}$
20	$y' = (x + y - 3)^2 - 2(x + y - 3)$		

1.4

Homogeneous first-order equations

If all terms in the coefficient functions $M(x,y)$ and $N(x,y)$ in the general first-order differential equation

$$M(x,y)\,dx = N(x,y)\,dy \tag{1}$$

are of the same degree in the variables x and y, then either of the substitutions $y = ux$ and $x = vy$ will reduce the equation to one which is separable

More generally, if $M(x,y)$ and $N(x,y)$ have the property that, for all positive values of λ, the substitution of λx for x and

λy for y converts them, respectively, into the expressions

$$\lambda^n M(x,y) \qquad \text{and} \qquad \lambda^n N(x,y)$$

then Eq. (1) can always be reduced to a separable equation by either of the substitutions $y = ux$ and $x = vy$

Functions with the property that the substitutions

$$x \longrightarrow \lambda x \qquad \text{and} \qquad y \longrightarrow \lambda y \qquad \lambda > 0$$

merely reproduce the original forms multiplied by λ^n are called **homogeneous functions of degree *n*** As a direct extension of this terminology, the differential equation (1) is said to be **homogeneous** when $M(x,y)$ and $N(x,y)$ are homogeneous functions *of the same degree.*

EXAMPLE 1

Is the function

$$F(x,y) = x(\ln\sqrt{x^2+y^2} - \ln y) + ye^{x/y}$$

homogeneous?

To decide this question, we replace x by λx and y by λy, getting

$$\begin{aligned} F(\lambda x,\lambda y) &= \lambda x(\ln\sqrt{\lambda^2x^2+\lambda^2y^2} - \ln \lambda y) + \lambda y e^{\lambda x/\lambda y} \\ &= \lambda x[(\ln\sqrt{x^2+y^2} + \ln\lambda) - (\ln y + \ln\lambda)] + \lambda y e^{x/y} \\ &= \lambda[x(\ln\sqrt{x^2+y^2} - \ln y) + ye^{x/y}] \\ &= \lambda F(x,y) \end{aligned}$$

The given function is, therefore, homogeneous of degree 1

If Eq. (1), assumed now to be homogeneous, is written in the form

$$\frac{dy}{dx} = \frac{M(x,y)}{N(x,y)}$$

it is evident that the fraction on the right is a homogeneous function of degree zero, since the same power of λ will multiply both numerator and denominator when the test substitutions $x \longrightarrow \lambda x$ and $y \longrightarrow \lambda y$ are made But if

$$\frac{M(\lambda x,\lambda y)}{N(\lambda x,\lambda y)} = \frac{M(x,y)}{N(x,y)}$$

it follows, assigning to the arbitrary symbol λ the value $1/x$ if x is positive and the value $-1/x$ if x is negative, that

$$\frac{M(x,y)}{N(x,y)} = \frac{M(\lambda x,\lambda y)}{N(\lambda x,\lambda y)} = \begin{cases} \dfrac{M(1,y/x)}{N(1,y/x)} & x > 0 \\ \dfrac{M(-1,-y/x)}{N(-1,-y/x)} & x < 0 \end{cases}$$

In either case it is clear that the result is a function of the fractional argument y/x. Thus, an alternative standard form for a homogeneous first-order differential equation is

(2) $$\frac{dy}{dx} = R\left(\frac{y}{x}\right)$$

Although in practice it is not necessary to reduce a homogeneous equation to the form (2) in order to solve it, the theory of the substitution $y = ux$ or $u = y/x$ is most easily developed when the equation is written in this form.

Now, if $y = ux$, then $dy/dx = u + x(du/dx)$ Hence, under this substitution, Eq (2) becomes

$$u + x\frac{du}{dx} = R(u)$$

or

$$x\,du = [R(u) - u]\,dx \tag{3}$$

If $R(u) \equiv u$, Eq (2) is simply

$$\frac{dy}{dx} = \frac{y}{x}$$

and this is separable at the outset If $R(u) \not\equiv u$, we can divide (3) by the product $x[R(u) - u]$, getting

$$\frac{du}{R(u) - u} = \frac{dx}{x}$$

The variables have now been separated, and the equation can be integrated at once Finally, by replacing u by its value y/x, we can obtain the equation defining y as a function of x

EXAMPLE 2

Solve the equation $(x^2 + 3y^2)\,dx - 2xy\,dy = 0$

By inspection, this equation is homogeneous, since all terms in the coefficient of each differential are of the second degree Hence, we substitute $y = ux$ and $dy = u\,dx + x\,du$, getting

$$(x^2 + 3u^2x^2)\,dx - 2x^2u(u\,dx + x\,du) = 0$$

or, dividing by x^2 and collecting terms,

$$(1 + u^2)\,dx - 2xu\,du = 0$$

Separating variables, we obtain

$$\frac{dx}{x} - \frac{2u\,du}{1 + u^2} = 0$$

and then, by integrating, we find

$$\ln|x| - \ln|1 + u^2| = c$$

This can be written as

$$\ln\left|\frac{x}{1 + u^2}\right| = \ln e^c = \ln k \qquad \text{where } k = e^c > 0$$

Hence, $|x/(1 + u^2)| = k$, or, replacing u by y/x and dropping absolute values,

$$\frac{x}{1 + (y/x)^2} = \pm k$$

Finally, clearing fractions, we have

$$x^3 = K(x^2 + y^2)$$

where, from the preceding steps, it appears that K can have any real value except zero. However, it is easy to verify by direct substitution that the function corresponding to $K = 0$, namely, $x = 0$, is also a solution of the given equation. Hence, in the general solution we have just obtained, K is actually unrestricted.

EXERCISES

1 Under what conditions, if any, do you think the substitution $x = vy$ would be more convenient than the substitution $y = ux$?

Find a general solution of each of the following equations.

2 $(x - 3y)\,dx = (3x + 2y)\,dy$

3 $(-x + 3y)\,dx = (x + y)\,dy$

4 $2x(dx + dy) + y(dy - 5\,dx) = 0$

5 $(x^2 + 2y^2)\,dx - xy\,dy = 0$

Find that particular solution of each of the following equations which satisfies the given conditions.

6 $xy\,dx = x^2\,dy - y^2\,dx$ $\qquad x = 1,\ y = 1$

7 $(3y^3 - x^3)\,dx = 3xy^2\,dy$ $\qquad x = 1,\ y = 2$

8 $(x + y)^2\,dx = xy\,dy$ $\qquad x = 1,\ y = 1$

9 $y\,dy = (2x + y)\,dx$ $\qquad x = 2,\ y = 1$

10 $x\,dy - y\,dx = \sqrt{x^2 + y^2}\,dx$ $\qquad x = 4,\ y = 3$

11 $(x^3 + y^3)\,dx = 2xy^2\,dy$ $\qquad x = 2,\ y = 1$

12 $\dfrac{dy}{dx} = \sec y/x + (y/x)$ $\qquad x = 2,\ y = \pi$

13 $(x^4 + y^4)\,dx = 2x^3y\,dy$ $\qquad x = 1,\ y = 0$

14 $(y^2 + 2xy)\,dx + 2x^2\,dy = 0$ $\qquad x = 1,\ y = -2$

15 If $aB \neq bA$, show that, by choosing d and D suitably, the equation

$$\frac{dy}{dx} = \frac{ax + by + c}{Ax + By + C}$$

can be reduced to a homogeneous equation by the substitutions

$$x = t + d \qquad \text{and} \qquad y = z + D$$

16 Discuss Exercise 15 in the case when $aB = bA$. (Hint: Recall Exercise 17, Sec. 1.3.)

Find a general solution of each of the following equations.

17 $y' = (x - y + 5)/(x + y - 1)$

18 $y' = (2x + 2y + 1)/(3x + y - 2)$

19 Give an example of a function which is homogeneous according to our definition but is not homogeneous if $f(\lambda x,\lambda y) = \lambda^n f(x,y)$ is required to hold for *all* real values of λ.

20 If $f(x,y)$ is a homogeneous function of degree n, show that

$$x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y} = nf$$

What is the generalization of this result to functions of more than two variables? (This is commonly referred to as **Euler's theorem for homogeneous functions**.)

1.5

Exact first-order equations

Associated with each suitably differentiable function of two variables $f(x,y)$ there is an expression called its **total differential,** namely,

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy$$

Conversely, if the differential equation

$$M(x,y)\,dx + N(x,y)\,dy = 0$$

has the property that

$$M(x,y) = \frac{\partial f}{\partial x} \qquad \text{and} \qquad N(x,y) = \frac{\partial f}{\partial y}$$

then it can be rewritten in the form

$$\frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy = df = 0$$

from which it follows that $f(x,y) = k$ is a solution for all values of the constant k. Equations of this sort are said to be **exact,** since, as they stand, their left members are *exact* differentials.

When $M(x,y)$ and $N(x,y)$ are sufficiently simple, it is possible to tell by inspection whether or not there exists a function f with the property that

$$\frac{\partial f}{\partial x} = M(x,y) \qquad \text{and} \qquad \frac{\partial f}{\partial y} = N(x,y)$$

In general, however, this cannot be done, and it is desirable to have a straightforward test to determine when a given first-order equation is exact. Such a criterion is provided by the following theorem.

THEOREM 1

If $\dfrac{\partial M}{\partial y}$ and $\dfrac{\partial N}{\partial x}$ are continuous, then the differential equation

$$M(x,y)\,dx + N(x,y)\,dy = 0$$

is exact if and only if $\dfrac{\partial M}{\partial y} = \dfrac{\partial N}{\partial x}$.

PROOF To prove the theorem, let us assume first that the given equation is exact. Under this assumption there exists a function f such that

$$M = \frac{\partial f}{\partial x} \qquad \text{and} \qquad N = \frac{\partial f}{\partial y}$$

Hence,

$$\frac{\partial M}{\partial y} = \frac{\partial^2 f}{\partial y\,\partial x} \qquad \text{and} \qquad \frac{\partial N}{\partial x} = \frac{\partial^2 f}{\partial x\,\partial y}$$

Moreover, from the familiar properties of partial derivatives, we know that, under our hypotheses,

$$\frac{\partial^2 f}{\partial y\,\partial x} = \frac{\partial^2 f}{\partial x\,\partial y}$$

Hence, $\dfrac{\partial M}{\partial y} = \dfrac{\partial N}{\partial x}$; and the "only if" part of the theorem is established.

To complete the proof we must now show that, if $\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}$, then there is a function f such that $\frac{\partial f}{\partial x} = M$ and $\frac{\partial f}{\partial y} = N$ To do this, let us first integrate $M(x,y)$ with respect to x, holding y fixed. This gives us the expression

$$(1) \qquad f(x,y) = \int_a^x M(x,y)\,dx + c(y) \qquad a \text{ arbitrary}$$

in which the integration "constant" is actually a function of y to be determined Clearly, $\frac{\partial f}{\partial x} = M(x,y)$, and our proof will be complete if we can determine $c(y)$ so that $\frac{\partial f}{\partial y} = N(x,y)$

Now, observing that, under our hypotheses, the operations of integrating with respect to x and differentiating with respect to y can legitimately be interchanged, and recalling our supposition that

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}$$

we have, from (1),

$$\begin{aligned}\frac{\partial f}{\partial y} &= \frac{\partial}{\partial y}\int_a^x M(x,y)\,dx + c'(y)\\ &= \int_a^x \frac{\partial M}{\partial y}\,dx + c'(y)\\ &= \int_a^x \frac{\partial N}{\partial x}\,dx + c'(y)\\ &= N(x,y) - N(a,y) + c'(y)\end{aligned}$$

Thus, $\frac{\partial f}{\partial y}$ will equal $N(x,y)$, as required, if $c(y)$ is determined so that

$$c'(y) = N(a,y)$$

that is, if

$$c(y) = \int_b^y N(a,y)\,dy \qquad b \text{ arbitrary}$$

We have thus shown that, if $\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}$, then

$$f(x,y) = \int_a^x M(x,y)\,dx + \int_b^y N(a,y)\,dy$$

is a function such that

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy = M(x,y)\,dx + N(x,y)\,dy$$

This establishes the "if" assertion of the theorem, and our proof is complete

COROLLARY 1

If the differential equation $M(x,y)\,dx + N(x,y)\,dy = 0$ is exact, then, for all values of k,

$$\int_a^x M(x,y)\,dx + \int_b^y N(a,y)\,dy = k$$

is a solution of the equation

EXAMPLE 1

Show that the equation $(2x + 3y - 2)\,dx + (3x - 4y + 1)\,dy = 0$ is exact, and find a general solution

Applying the test provided by Theorem 1, we find

$$\frac{\partial M}{\partial y} = \frac{\partial(2x + 3y - 2)}{\partial y} = 3 \quad \text{and} \quad \frac{\partial N}{\partial x} = \frac{\partial(3x - 4y + 1)}{\partial x} = 3$$

Since the two partial derivatives are equal, the equation is exact Its solution can, therefore, be found by means of Corollary 1, Theorem 1

$$\int_a^x (2x + 3y - 2)\,dx + \int_b^y (3a - 4y + 1)\,dy = k$$

$$(x^2 + 3xy - 2x)\Big|_a^x + (3ay - 2y^2 + y)\Big|_b^y = k$$

$$x^2 + 3xy - 2y^2 - 2x + y = k + a^2 + 3ab - 2b^2 - 2a + b = K$$

Occasionally an equation which is not exact can be made exact by multiplying it by some simple expression In fact, it can be shown* that every first-order equation which possesses a general solution can be made exact by multiplying it by a suitable factor, called an **integrating factor.** In general, the determination of an integrating factor for a given equation is very difficult However, as the following examples show, in particular cases an integrating factor can often be found by inspection

EXAMPLE 2

Show that $1/(x^2 + y^2)$ is an integrating factor for the equation $(x^2 + y^2 - x)\,dx - y\,dy = 0$

If the given equation is multiplied by the indicated factor, it can be rewritten in the form

$$\left(1 - \frac{x}{x^2 + y^2}\right)dx - \frac{y}{x^2 + y^2}\,dy = 0$$

The test provided by Theorem 1 can be used to show that this equation is exact, and Corollary 1, Theorem 1, can be used to obtain the solution However, it is simpler to observe that the last equation can also be written

$$dx - \frac{x\,dx + y\,dy}{x^2 + y^2} = 0 \quad \text{or} \quad dx - \tfrac{1}{2}d[\ln(x^2 + y^2)] = 0$$

Hence, integrating,

$$x - \ln\sqrt{x^2 + y^2} = k$$

EXAMPLE 3

Find an integrating factor for the equation $y\,dx + (x^2y^3 + x)\,dy = 0$, and solve the equation

Since this equation can be rewritten in the form

$$(y\,dx + x\,dy) + x^2y^3\,dy = 0$$

and since $y\,dx + x\,dy = d(xy)$, it is natural to multiply the equation by $1/x^2y^2$, getting

$$\frac{d(xy)}{x^2y^2} + y\,dy = 0$$

---

* See, for instance, Golomb and Shanks, *op cit*, pp 52–53.

This equation can now be integrated by inspection, and we have

$$-\frac{1}{xy}+\frac{y^2}{2}=k$$

EXAMPLE 4

Find an integrating factor for the equation $x\,dy - y\,dx = (4x^2 + y^2)\,dy$, and solve the equation

In this equation, the terms on the left seem related to

$$d\left(\frac{y}{x}\right)=\frac{x\,dy-y\,dx}{x^2}\qquad\text{or, equally well,}\qquad d\left(\frac{x}{y}\right)=\frac{y\,dx-x\,dy}{y^2}$$

If we pursue the first suggestion and multiply the equation by $1/x^2$, we obtain

$$d\left(\frac{y}{x}\right)=\left(4+\frac{y^2}{x^2}\right)dy$$

This equation is still not exact, but it is separable, and division by $4 + y^2/x^2$ gives us

$$\frac{d(y/x)}{4+(y/x)^2}=dy$$

Integrating this, we have finally

$$\frac{1}{2}\operatorname{Tan}^{-1}\left(\frac{y}{2x}\right)=y+k$$

The results of the last three examples suggest the following observations, which are often helpful

a If a first-order differential equation contains the combination $x\,dx + y\,dy$, try some function of $x^2 + y^2$ as an integrating factor

b If a first-order differential equation contains the combination $y\,dx + x\,dy$, try some function of xy as an integrating factor

c If a first-order differential equation contains the combination $x\,dy - y\,dx$, try $1/x^2$ or $1/y^2$ as an integrating factor

EXERCISES

Find a general solution of each of the following equations

1 $(3x^2 - 6xy)\,dx - (3x^2 + 2y)\,dy = 0$

2 $(y^2 - 1)\,dx + (2xy - \sin y)\,dy = 0$

3 $(2xy + x^3)\,dx + (x^2 + y^2)\,dy = 0$

4 $(x\sqrt{x^2+y^2} + y)\,dx + (y\sqrt{x^2+y^2} + x)\,dy = 0$

5 $y(1 + xy)\,dx - (x - 2y)\,dy = 0$

6 $3(y^4 + 1)\,dx + 4xy^3\,dy = 0$

7 $(xy^2 - y)\,dx + x(xy - 1)\,dy = 0$

8 $y\,dx - dy = x^2y^2\,dx + x\,dy$

9 $2y\,dx + 3x\,dy = dx/xy^3 - dy/y^4$

Solve each of the following equations by two methods

10 $2y\,dx + (3y - 2x)\,dy = 0$

11 $(x + y)\,dx + (x - y)\,dy = 0$

12 $\sqrt{x^2+y^2}\,dx = x\,dy - y\,dx$

13 $x\,dy + y\,dx = dx/y - dy/x$

14 Show that the arbitrary constants a and b which appear in the formula of Corollary 1, Theorem 1, add no generality to the solution (Hint Consider the partial derivatives with respect to a and b of the left-hand member of the formula)

15 If ϕ is an integrating factor of the equation $M(x,y)\,dx + N(x,y)\,dy = 0$, show that ϕ satisfies the partial differential equation

$$M\frac{\partial \phi}{\partial y} - N\frac{\partial \phi}{\partial x} + \phi\left(\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}\right) = 0$$

1.6

Linear first-order equations

By definition, a linear first-order differential equation cannot contain products, powers, or other nonlinear combinations of y or y' Hence, its most general form is

$$F(x)\frac{dy}{dx} + G(x)y = H(x)$$

If we divide this equation by $F(x)$ and rename the coefficients, it appears in the more usual form

$$\frac{dy}{dx} + P(x)y = Q(x) \tag{1}$$

The presence of two terms on the left side of (1) involving, respectively, dy/dx and y suggests strongly that this expression is in some way related to the derivative of a product, say $\phi(x)y$, having y as one factor Now the derivative of $\phi(x)y$ is

$$\phi(x)\frac{dy}{dx} + \frac{d\phi(x)}{dx}y \tag{2}$$

and the left member of (1) can be made identically equal to this, provided we first multiply Eq (1) by $\phi(x)$, getting

$$\phi(x)\frac{dy}{dx} + \phi(x)P(x)y = \phi(x)Q(x) \tag{3}$$

and then make the second terms in (2) and (3) equal by choosing $\phi(x)$ such that

$$\frac{d\phi(x)}{dx} = \phi(x)P(x)$$

This is a simple separable equation, any nontrivial solution of which will meet our requirements Hence, we can write, in particular,

$$\frac{d\phi(x)}{\phi(x)} = P(x)\,dx$$

$$\ln|\phi(x)| = \int P(x)\,dx$$

$$\phi(x) = \exp\left[\int P(x)\,dx\right]\dagger$$

Thus, after Eq (1) is multiplied by the factor

$$\phi(x) = \exp\left[\int P(x)\,dx\right]$$

---

† The notation $\exp[f(x)]$ is frequently used in place of $e^{f(x)}$, especially when $f(x)$ is a complicated expression

it can be written in the form

$$\frac{d}{dx}\left(\exp\left[\int P(x)\,dx\right]y\right) = Q(x)\exp\left[\int P(x)\,dx\right]$$

Now, the left-hand side is an exact derivative and, hence, can be integrated at once Moreover, the right-hand side is a function of x only and can, therefore, be integrated also, with at most practical difficulties requiring numerical integration Thus we have, on performing these integrations,

$$y\exp\left[\int P(x)\,dx\right] = \int Q(x)\exp\left[\int P(x)\,dx\right]dx + c$$

and finally

$$y = \exp\left[-\int P(x)\,dx\right]\int Q(x)\exp\left[\int P(x)\,dx\right]dx + c\exp\left[-\int P(x)\,dx\right] \tag{4}$$

The factor $\exp\left[\int P(x)\,dx\right]$ which converts the general linear differential equation (1) into one which can be integrated by inspection is, of course, another example of an integrating factor

Equation (4) should *not* be remembered as a formula for the solution of Eq. (1) A linear first-order equation should be solved by actually carrying out the steps we have described

a Compute the integrating factor $\exp\left[\int P(x)\,dx\right]$
b Multiply the given equation by this factor
c Integrate both sides of the resulting equation, taking advantage of the fact that the integral of the left member is *always* just y times the integrating factor.
d Solve the integrated equation for y

EXAMPLE 1

Find the solution of the equation $(1 + x^2)(dy - dx) = 2xy\,dx$ for which $y = 1$ when $x = 0$

Dividing the given equation by $(1 + x^2)\,dx$ and transposing, we have

$$\frac{dy}{dx} - \frac{2x}{1 + x^2}y = 1 \tag{5}$$

which is a linear first-order equation In this case $P(x) = -2x/(1 + x^2)$, thus, the integrating factor is

$$\exp\left(\int \frac{-2x}{1 + x^2}\,dx\right) = e^{-\ln(1+x^2)} = e^{\ln(1+x^2)^{-1}} = \frac{1}{1 + x^2}\dagger$$

Multiplying Eq (5) by this factor gives the equation

$$\frac{1}{1 + x^2}\frac{dy}{dx} - \frac{2x}{(1 + x^2)^2}y = \frac{1}{1 + x^2}$$

Integrating this, remembering that the integral of the left member is just y times the integrating factor, we have

$$\frac{y}{1 + x^2} = \operatorname{Tan}^{-1}x + k \qquad \text{or} \qquad y = (1 + x^2)\operatorname{Tan}^{-1}x + k(1 + x^2)$$

---

† Note that $e^{\ln u} = u$ for any expression u

To find the specific solution required, we substitute the given conditions $x = 0$, $y = 1$ into the general solution, getting $1 = 0 + k$. The required solution is, therefore,

$$y = (1 + x^2) \operatorname{Tan}^{-1} x + (1 + x^2)$$

EXERCISES

Find a general solution of each of the following equations

1 $x^2\,dy + (2xy - x + 1)\,dx = 0$
2 $x\,dy = (2y + 3x^4 + x^2)\,dx$
3 $xy' + (1 + x)y = e^{-x}$
4 $(2y + x^2)\,dx = x\,dy$
5 $y' + y/(1 - x) = x^2 - x$
6 $y' = 2y/(x + 1) + (x + 1)^3$
7 $y' + y \cot x = \sin 2x$
8 $y' + 2xy + x = e^{-x^2}$
9 $2x\,dy = (2x - y)\,dx$
10 $y' + y \tan x = \sec x$
11 $y^2\,dx + (3xy - 4y^3)\,dy$ (Hint Consider x as the dependent variable)
12 $y'' + y'/(x - 1) = x - 1$ (Hint Consider y' as the dependent variable)

Find the particular solution of each of the following equations which satisfies the indicated conditions

13 $(1 + x^2)\,dy = (1 + xy)\,dx$ $x = 0,\ y = 1$
14 $y' - e^{-x^2} + (1 + 2x)y = 0$ $x = 0,\ y = 2$
15 $x\,dy + (x^2 - 3y)\,dx = 0$ $x = -1,\ y = 1$ *and* $x = 1,\ y = -1$
16 Show that the substitution $z = y^{1-n}$ will reduce the **Bernoulli equation***

$$y' + P(x)y = Q(x)y^n$$

to a linear first-order equation

Find a general solution of each of the following equations

17 $3xy' + y + x^2y^4 = 0$
18 $dy = (xy^2 + 3xy)\,dx$
19 $y' + y = x/y$

20 Prove that no extra generality in the final answer results from using $e^{\int P(x)dx+c}$ instead of just $e^{\int P(x)dx}$ as an integrating factor for the equation $y' + P(x)y = Q(x)$

1.7

Applications of first-order differential equations

The mathematical formulation of physical problems involving continuously changing quantities often leads to differential equations of the first order. The following examples will make clear how such equations arise and how they are handled

EXAMPLE 1

A tank is initially filled with 100 gal of salt solution containing 1 lb of salt per gallon Fresh brine containing 2 lb of salt per gallon runs into the tank at the rate of 5 gal/min, and the mixture, assumed to be kept uniform by stirring, runs out at the same rate Find the amount of salt in the tank at any time t, and determine how long it will take for this amount to reach 150 lb

Let Q lb be the total amount of salt in solution in the tank at any time t, and let dQ be the increase in this amount during the infinitesimal interval of time dt At any time t, the amount of salt per gallon of solution is therefore $Q/100$ (lb/gal) Now the change dQ in the total amount of salt in the tank is clearly the net gain in the interval dt due to the fresh brine running into and the mixture running out of the tank The rate at which salt enters the tank is

---

* Named for the Swiss mathematician James Bernoulli (1654–1705), a member of a family which in little more than a century produced eight distinguished mathematicians

$$5\ (\text{gal/min}) \times 2\ (\text{lb/gal}) = 10\ (\text{lb/min})$$

Hence, in the interval dt the gain in salt from this source is

$$10\ (\text{lb/min}) \times dt\ (\text{min}) = 10\,dt\ (\text{lb})$$

Likewise, since the concentration of salt in the mixture as it leaves the tank is the same as the concentration $Q/100$ in the tank itself, the amount of salt leaving the tank in the interval dt is

$$5\ (\text{gal/min}) \times \frac{Q}{100}\ (\text{lb/gal}) \times dt\ (\text{min}) = \frac{Q}{20}\,dt\ (\text{lb})$$

Therefore, $$dQ = \left(10 - \frac{Q}{20}\right)dt$$

This equation can be written in the form

$$(1) \qquad \frac{dQ}{200 - Q} = \frac{dt}{20}$$

and handled as a separable equation, or it can be written

$$(2) \qquad \frac{dQ}{dt} + \frac{Q}{20} = 10$$

and treated as a linear equation.

Considering it as a linear equation, we must first compute the integrating factor

$$e^{\int P\,dt} = e^{\int dt/20} = e^{t/20}$$

Multiplying Eq (2) by this factor gives

$$e^{t/20}\left(\frac{dQ}{dt} + \frac{Q}{20}\right) = 10e^{t/20}$$

From this, by integration, we obtain

$$Qe^{t/20} = 200e^{t/20} + k \qquad \text{or} \qquad Q = 200 + ke^{-t/20}$$

Substituting the initial conditions $t = 0$, $Q = 100$, we find

$$100 = 200 + k \qquad \text{or} \qquad k = -100$$

Hence, $$Q = 200 - 100e^{-t/20}$$

To find how long it will be before there is 150 lb of salt in the tank, we must find the value of t such that

$$150 = 200 - 100e^{-t/20} \qquad \text{or} \qquad e^{-t/20} = \tfrac{1}{2}$$

From this we have at once

$$-\frac{t}{20} = \ln\frac{1}{2} = -\ln 2 = -0\,693 \qquad \text{and} \qquad t = 13\,9\ \text{min}$$

EXAMPLE 2

A hemispherical tank of radius R is initially filled with water At the bottom of the tank there is a hole of radius r through which the water drains under the influence of gravity Find the depth of the water at any time t, and determine how long it will take the tank to drain compl ely.

Let the origin be chosen at the lowest point of the tank, let y be the instantaneous depth of the water, and let x be the instantaneous radius of the free surface of the water (Fig 1.3). Then in the infinitesimal interval dt the water level will fall by the amount dy, and the resultant decrease in the volume of water in the tank will be

$$dV = \pi x^2\, dy$$

This, of course, must equal the volume of water that leaves the orifice during the time dt Now from **Torricelli's law,*** the velocity with which a liquid issues from an orifice is

$$v = \sqrt{2gh}$$

where g is the acceleration of gravity and y is the instantaneous height, or **head,** of the liquid above the orifice In the interval dt, then, a stream of water of length $\sqrt{2gy}\,dt$ and of cross-section area πr^2† will emerge from the outlet The volume of this amount of water is

$$dV = \pi r^2 \sqrt{2gy}\, dt$$

Hence, equating the two expressions for dV, we obtain the differential equation

$$\pi x^2\, dy = -\pi r^2 \sqrt{2gy}\, dt \tag{3}$$

the minus sign indicating that as t increases, the depth y decreases

Before this equation can be solved, it is necessary that x be expressed in terms of y This is easily done through the use of the equation of the circle which describes the vertical cross section of the tank

$$x^2 + (y - R)^2 = R^2 \qquad \text{or} \qquad x^2 = 2yR - y^2$$

Using this, the differential equation (3) can be written

$$\pi(2yR - y^2)\, dy = -\pi r^2 \sqrt{2gy}\, dt$$

This is a simple separable equation which can be solved without difficulty

$$(2Ry^{1/2} - y^{3/2})\, dy = -r^2 \sqrt{2g}\, dt$$

$$\tfrac{4}{3}Ry^{3/2} - \tfrac{2}{5}y^{5/2} = -r^2 \sqrt{2g}\, t + c$$

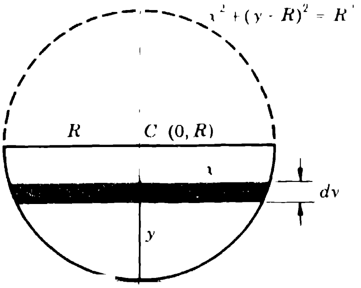


FIGURE 1.3
A vertical plane section through the center of a hemispherical tank

---

* Named for the Italian mathematician and physicist Evangelista Torricelli (1608–1647)

† This neglects the fact that the stream contracts near the orifice How much the cross section of the stream decreases depends in a very complicated way upon the size and shape of both the tank and the orifice and also upon the head However, in most practical problems reasonably accurate answers can be obtained by assuming that the cross section of the stream just after it leaves the orifice is 0 6 times the area of the orifice

Since $y = R$ when $t = 0$, we find

$$\tfrac{14}{15}R^{5/2} = c$$

and thus
$$\tfrac{4}{3}Ry^{3/2} - \tfrac{2}{5}y^{5/2} = -r^2\sqrt{2g}\,t + \tfrac{14}{15}R^{5/2}$$

This is the equation which expresses the instantaneous depth y as a function of t

To find how long it will take the tank to empty, we must determine the value of t corresponding to $y = 0$

$$0 = -r^2\sqrt{2g}\,t + \tfrac{14}{15}R^{5/2}$$

$$t = \frac{14}{15}\frac{R^{5/2}}{r^2\sqrt{2g}}$$

EXAMPLE 3

The rate at which a solid substance dissolves varies directly as the amount of undissolved solid present in the solvent and as the difference between the instantaneous concentration and the saturation concentration of the substance Twenty pounds of solute is dumped into a tank containing 120 lb of solvent, and at the end of 12 min the concentration is observed to be 1 part in 30. Find the amount of solute in solution at any time t if the saturation concentration is 1 part of solute to 3 parts of solvent

If Q is the amount of the material in solution at time t, then $20 - Q$ is the amount of undissolved material at that time and $Q/120$ is the corresponding concentration Hence, according to the given law,

$$\frac{dQ}{dt} = k(20 - Q)\left(\frac{1}{3} - \frac{Q}{120}\right) = \frac{k}{120}(20 - Q)(40 - Q)$$

This is a simple separable equation, and we have at once

$$\frac{dQ}{(20 - Q)(40 - Q)} = \frac{k}{120}\,dt$$

To integrate the left member it is convenient to use the method of partial fractions and write

$$\frac{1}{(20 - Q)(40 - Q)} = \frac{A}{20 - Q} + \frac{B}{40 - Q} = \frac{A(40 - Q) + B(20 - Q)}{(20 - Q)(40 - Q)}$$

This will be an identity if and only if

$$1 = A(40 - Q) + B(20 - Q)$$

Setting $Q = 20$ and $Q = 40$, in turn, we find from this that

$$A = \tfrac{1}{20} \qquad \text{and} \qquad B = -\tfrac{1}{20}$$

Hence the differential equation can be written

$$\frac{1}{20}\left(\frac{1}{20 - Q} - \frac{1}{40 - Q}\right)dQ = \frac{k}{120}\,dt$$

and, integrating, we have

$$(4) \qquad -\ln(20 - Q) + \ln(40 - Q) = \frac{k}{6}t + c$$

When $t = 0$, the amount $Q = Q_0$ of dissolved material is zero Hence

$$-\ln 20 + \ln 40 = c \qquad \text{or} \qquad c = \ln 2$$

and Eq. (4) can be written

$$(5)\qquad \ln\frac{40-Q}{2(20-Q)} = \frac{k}{6}t$$

To find k we use the fact that when $t = 12$, the concentration $Q/120$ is $\frac{1}{30}$, or $Q = 4$. Hence, substituting these values,

$$\ln \tfrac{36}{32} = 2k \qquad \text{or} \qquad k = \tfrac{1}{2}\ln \tfrac{9}{8} = 0.05889$$

Passing to exponential form from Eq. (5), in order to solve for Q, we have

$$\frac{40-Q}{40-2Q} = e^{0.0098t}$$

and finally

$$Q = \frac{40 - 40e^{0.0098t}}{1 - 2e^{0.0098t}} = \frac{40(1 - e^{-0.0098t})}{2 - e^{-0.0098t}}$$

EXERCISES

1. Under certain conditions it is observed that the rate at which atmospheric pressure changes with altitude is proportional to the pressure. If the pressure is 14.7 lb/in.² at sea level and if it has fallen to one-half this value at 18,000 ft, find the formula for the pressure at any height.
2. Although water is often assumed to be incompressible, it actually is not. In fact, using pounds and feet as units, the weight of a cubic foot of water under pressure p is approximately $w(1 + kp)$, where $w = 64$, $k = 2 \times 10^{-8}$, and p is measured from standard atmospheric pressure as an origin. Using this information, find the pressure at any depth y below the surface of the ocean. At a depth of 6 miles, by what factor does the actual pressure exceed the pressure computed on the assumption that water is incompressible?
3. Radium disintegrates at a rate proportional to the amount of radium instantaneously present. If one-half of any given amount of radium will disappear in 1,590 years, what fraction will disintegrate during the first century? during the tenth century?
4. According to **Lambert's law of absorption**,[*] when light passes through a transparent medium, the amount absorbed by any thin layer of the material is proportional to the amount incident on that layer and to the thickness of the layer. In his deep-sea explorations off Bermuda, Beebe observed that at a depth of 50 ft the intensity of illumination was 10 candles/ft², and that at 250 ft it had fallen to 0.2 candle/ft². Find the law connecting intensity with depth in this case.
5. It is a fact of common experience that, when a rope is wound around a rough cylinder, a small force at one end can resist a much larger force at the other. Quantitatively, it is found that, throughout the portion of the rope in contact with the cylinder, the change in tension per unit length is proportional to the tension, the proportionality constant being the coefficient of friction between the rope and the cylinder divided by the radius of the cylinder. Assuming a coefficient of friction of 0.35, how many times must a rope be snubbed around a post 1 ft in diameter in order that a man holding one end can resist a force 200 times greater than he can exert?
6. When ethyl acetate in dilute aqueous solution is heated in the presence of a small amount of acid, it decomposes according to the following equation

$$\underset{\text{ethyl acetate}}{CH_3COOC_2H_5} + \underset{\text{water}}{H_2O} \rightarrow \underset{\text{acetic acid}}{CH_3COOH} + \underset{\text{ethyl alcohol}}{C_2H_5OH}$$

---

[*] Named for the German mathematician and astronomer Johann Heinrich Lambert (1728–1777).

Since this reaction takes place in dilute solution, the quantity of water present is so great that the loss of the small amount which combines with the ethyl acetate produces no appreciable change in the total amount. Hence, of the reacting substances only the ethyl acetate suffers a measurable change in concentration. A chemical reaction of this sort, in which the concentration of only one reacting substance changes, is called a **first-order reaction.** It is a law of physical chemistry that the rate at which a substance is being used up, i.e., transformed, in a first-order reaction is proportional to the amount of that substance instantaneously present. If the initial concentration of ethyl acetate is C_0, find the expression for its instantaneous concentration at any time t.

7 In some chemical reactions where two substances combine to form a third, the amount of each of the reacting substances changes appreciably. In such cases it is observed that the rate at which the resulting compound is formed is proportional to the product of the untransformed amounts of the two reacting substances. If two substances combine in the ratio 1:2, by weight, to form a third substance, and if it is observed that, 10 min after 10 grams of the first substance and 20 grams of the second are mixed, the amount of the product which has been formed is 5 grams, find an expression for the amount of the product present at any time.

8 Work Exercise 7 given that, instead of 10 grams of the first substance and 20 grams of the second, 20 grams of each substance are mixed.

9 A mothball loses mass by evaporation at a rate proportional to its instantaneous surface area. If half its mass is lost in 100 days, how long will it take its radius to decrease to half its initial value? How long will it be before the mothball disappears completely?

10 When a volatile substance is placed in a sealed container, molecules leave its surface at a rate proportional to the area of the surface and return at a rate proportional to the amount which has evaporated. If a volatile material is spread evenly to a depth h over the bottom of a closed box, find the depth of the material at any time. Under what conditions, if any, will all the material eventually evaporate?

11 A rapidly rotating flywheel, after power is shut off, "coasts" to rest under the retarding influence of a friction torque which is proportional to the instantaneous angular velocity ω. If the moment of inertia of the flywheel is I and if its initial velocity is ω_0, find its instantaneous angular velocity as a function of time. How long will it take the flywheel to come to rest? (Hint: Use **Newton's law in torsional form,**

Moment of inertia × *angular acceleration* = *torque*

to set up the differential equation describing the motion.)

12 The friction torque acting to slow down a flywheel is actually not proportional to the first power of the angular velocity at all speeds. As a more realistic example than Exercise 11, suppose that a flywheel of moment of inertia $I = 7.5$ lb-ft sec² coasts to rest from an initial speed of 1,000 rad/min under the influence of a retarding torque T estimated to be the following:

$$T = \begin{cases} \dfrac{\sqrt{\omega}}{10}\ \text{ft-lb} & 0 < \omega < 100\ \text{rad/min} \\[2ex] \dfrac{1}{10}\left(7.5 + \dfrac{\omega^2}{4{,}000}\right)\text{ft-lb} & 100 < \omega < 1{,}000\ \text{rad/sec} \end{cases}$$

Find ω as a function of t, and determine how long it will take the flywheel to come to rest.

13 A body weighing w lb falls from rest under the influence of gravity and a retarding force due to air resistance, assumed to be proportional to the velocity. Find the equations expressing the velocity of fall and the distance fallen, as functions of t, and verify that these reduce to the ideal laws

$$v = gt \qquad \text{and} \qquad s = \tfrac{1}{2}gt^2$$

when the coefficient of air resistance approaches zero. (Hint. Use **Newton's law,**

$$\textit{Mass} \times \textit{acceleration} = \textit{force}$$

to set up the differential equation which describes the motion.)

14 Work Exercise 13, given that the retarding force due to air resistance is proportional to the square of the velocity of fall.

15 A body falls from rest from a height so great that the fact that the force of gravity varies inversely as the square of the distance from the center of the earth cannot be neglected. Find the equations expressing the velocity of fall and the distance fallen as functions of t in the ideal case in which air resistance is neglected. [Hint. $dv/dt = (dv/dy)(dy/dt) = v(dv/dy)$.]

16 Under the conditions of Exercise 15, determine the minimum initial velocity with which a body must be projected upward if it is to leave the earth and never return.

17 A particle of mass m moves along the x-axis under the influence of a force which is directed toward the origin and proportional to the distance of the particle from the origin. If the body starts from rest at the point where $x = x_0$, find the equations that express its velocity and its distance from the origin as functions of t. (Note the hint to Exercise 15.)

18 A tank contains 100 gal of brine in which 50 lb of salt is dissolved. Brine containing 2 lb/gal of salt runs into the tank at the rate of 3 gal/min, and the mixture, assumed to be kept uniform by stirring, runs out of the tank at the rate of 2 gal/min. Assuming the tank sufficiently large to avoid overflow, find the amount of salt in the tank as a function of t.

19 Work Exercise 18 with the rates of influx and efflux interchanged.

20 Work Example 3 given that the saturation concentration is 1 part of solute to 12 parts of solvent.

21 Work Example 3 given that the saturation concentration is 1 part of solute to 6 parts of solvent.

22 Work Example 3, with *concentration* defined as the ratio of solute to solution instead of solute to solvent.

23 According to **Newton's law of cooling,** the rate at which the temperature of a body decreases is proportional to the difference between the instantaneous temperature of the body and the temperature of the surrounding medium. If a body whose temperature is initially 100°C is allowed to cool in air which remains at the constant temperature 20°C, and if it is observed that in 10 min the body has cooled to 60°, find the temperature of the body as a function of time.

24 A tank and its contents weigh 100 lb. The average heat capacity of the system is 0.5 Btu/(lb)(°F). The liquid in the tank is heated by an immersion heater which delivers 100 Btu/min. Heat is lost from the system at a rate proportional to the difference between the temperature of the system, assumed constant throughout at any instant, and the temperature of the surrounding air, the proportionality constant being 2 Btu/(min)(°F). If the air temperature remains constant at 70° and if the initial temperature of the tank and its contents is 55°, find the temperature of the tank at any time.

25 According to **Fourier's law of heat conduction,** the amount of heat in Btu per unit time flowing through an area is proportional to the area and to the temperature gradient, in degrees per unit length, normal to the area. On the basis of this law, obtain a formula for the amount of heat lost per unit time from 1 ft^2 of furnace wall h ft thick, if the temperature in the furnace is T_0 and if the air temperature outside the furnace is T_1. What is the temperature distribution through the furnace wall?

26 Using Fourier's law of heat conduction, obtain a formula for the amount of heat lost per unit time from l ft of pipe of radius r_0 carrying steam at temperature T_0 if the pipe is covered with w in. of insulation, the outer surface of which remains at the constant temperature T_1. What is the temperature distribution through the insulation?

27 The inner and outer surfaces of a hollow sphere are maintained at the respective temperatures T_0 and T_1. If the inner and outer radii of the spherical shell are r_0 and r_1, find the amount of heat lost from the sphere per unit time What is the temperature distribution through the shell?

28 When a condenser of capacity C is being charged through a resistance R by a battery which supplies a constant voltage E, the instantaneous charge Q on the condenser satisfies the differential equation

$$R\frac{dQ}{dt} + \frac{Q}{C} = E$$

Find Q as a function of t if the condenser is initially uncharged How long will it be before the condenser is half charged?

29 When a switch is closed in a circuit containing a resistance R, an inductance L, and a battery which supplies a constant voltage E, the current i builds up at a rate defined by the relation

$$L\frac{di}{dt} + Ri = E$$

Find i as a function of t How long will it take i to reach one-half of its final value?

30 In Exercise 28, find Q as a function of t if the battery is replaced by a generator which supplies an alternating voltage equal to $E_0 \sin \omega t$

31 In Exercise 29, find i as a function of t if the battery is replaced by a generator which supplies an alternating voltage equal to $E_0 \cos \omega t$ What is the phase difference between the impressed voltage and the resultant current after the current has been flowing a long time?

32 A vertical cylindrical tank of radius r is filled with liquid to a depth h When the tank is rotated about its axis, centrifugal force tends to drive the liquid outward from the center of the tank Under steady conditions of rotation with constant angular velocity ω, find the equation of the curve in which the free surface of the liquid is intersected by a plane through the axis of the cylinder, assuming the tank to be sufficiently deep that no liquid is spilled over the edge

33 A weight W is to be supported by a column having the shape of a solid of revolution If the material of the column weighs ρ lb/ft^3, and if the radius of the upper base of the column is to be r_0, determine how the radius of the column should vary in order that at all cross sections the load per unit area will be the same

34 Work Example 2 if the tank has the shape of an inverted right circular cone of radius R and height h.

35 Work Example 2 if the tank has the shape of a vertical right circular cylinder of radius R and height h and if, in addition to a hole of radius r in the bottom, there is also a hole of radius r in the side at a distance of $h/2$ above the base

36 A cylindrical tank is l ft long and has semicircular end sections of radius r ft The tank is placed with its axis horizontal and is initially filled with water How long will it take the tank to drain through a hole of area a ft^2 in the bottom of the tank?

37 A vertical cylindrical tank of radius r and height h has a narrow crack of width w running vertically from top to bottom If the tank is initially filled with water and allowed to drain through the crack under the influence of gravity, find the instantaneous depth of the water in the tank as a function of t How long will it take the tank to empty? (Hint First imagine the crack to be a series of adjacent orifices, and integrate to find the total efflux from the crack in the infinitesimal interval dt)

38 Water flows into a vertical cylindrical tank of cross-section area A ft^2 at the rate of Q ft^3/min At the same time the water flows out under the influence of gravity through a hole of area a ft^2 in the base of the tank If the water is initially h ft deep, find the instantaneous depth as a function of t.

39 If two families of curves have the property that each member of either family cuts every member of the other family at right angles, the curves of either family are said t be

orthogonal trajectories of the curves of the other family. Find the orthogonal trajectories of the curves of the family $2x^2 + y^2 = kx$ [Hint Show that, at a general point (x,y), the slope of that curve of the given family which passes through that point is given by the formula

$$y' = \frac{-2x^2 + y^2}{2xy}$$

and then find the curves whose slopes are given by the negative reciprocal of this expression]

40 Find the orthogonal trajectories of the curves of the family

$$y^2 - x^2 = kx$$

CHAPTER TWO

Linear Differential Equations with Constant Coefficients

2.1

The general linear second-order equation

The general linear differential equation of the second order can be written in the standard form

$$y'' + P(x)y' + Q(x)y = R(x) \tag{1}$$

where P, Q, and R are known functions. Clearly, no loss of generality results from taking the coefficient of y'' to be unity, since this can always be accomplished by division. Because of the presence of the term $R(x)$, which is unlike the other terms in that it does not contain the dependent variable y or any of its derivatives, Eq. (1) is said to be **nonhomogeneous.** If $R(x)$ is identically zero, we have the so-called **homogeneous** equation*

$$y'' + P(x)y' + Q(x)y = 0 \tag{2}$$

In general, neither Eq. (1) nor Eq. (2) can be solved in terms of known functions. The theory associated with such special cases as have been studied at length is, for the most part, very difficult. At this stage we shall consider in detail only the simple, though highly important, case in which $P(x)$ and $Q(x)$ are constants. However, both as an illustration of how certain properties of the solutions of a differential equation can be established even though the form of those solutions is unknown and also because we shall have need of the results themselves, we shall begin by proving three fundamental theorems pertaining to the solutions of the general equations (1) and (2).

---

* It is regrettable that in describing linear equations of all orders the word *homogeneous* should be used in a manner totally unlike its use in describing equations of the first order (Sec. 1.4). The usage is universal, however, and must be accepted.

THEOREM 1

If y_1 and y_2 are any solutions of the homogeneous equation

$$y'' + P(x)y' + Q(x)y = 0$$

then $y_3 = c_1y_1 + c_2y_2$, where c_1 and c_2 are arbitrary constants, is also a solution.

PROOF To establish this theorem, it is necessary only to substitute the expression for y_3 into the given differential equation and verify that it is satisfied

$$\begin{aligned} y_3'' + P(x)y_3' + Q(x)y_3 &= (c_1y_1 + c_2y_2)'' + P(x)(c_1y_1 + c_2y_2)' \\ &\qquad + Q(x)(c_1y_1 + c_2y_2) \\ &= (c_1y_1'' + c_2y_2'') + P(x)(c_1y_1' + c_2y_2') \\ &\qquad + Q(x)(c_1y_1 + c_2y_2) \\ &= [y_1'' + P(x)y_1' + Q(x)y_1]c_1 \\ &\qquad + [y_2'' + P(x)y_2' + Q(x)y_2]c_2 \\ &= 0 \quad c_1 + 0 \quad c_2 = 0 \end{aligned}$$

where the coefficients of c_1 and c_2 vanish identically because, by hypothesis, both y_1 and y_2 are solutions of the homogeneous equation (2)

Theorem 1 assures us that, if we have two solutions of the homogeneous equation (2), then we can obtain infinitely many other solutions simply by forming arbitrary linear combinations of these two However, it leaves completely unanswered the important question of whether or not *all* solutions of (2) can be obtained from the pair (y_1,y_2) in this fashion To decide this point we need the stronger result contained in the next theorem

THEOREM 2

If y_1 and y_2 are two solutions of the homogeneous equation

$$y'' + P(x)y' + Q(x)y = 0$$

for which

$$W(y_1,y_2)\dagger \equiv \begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix} = y_1y_2' - y_2y_1' \neq 0$$

and if $\int P(x)\,dx$ exists, then there exist constants c_1 and c_2 such that any solution y_3 of the homogeneous equation can be expressed in the form $y_3 = c_1y_1 + c_2y_2$

PROOF To prove this theorem, it is convenient to show first that any pair of solutions of Eq (2), say y_i and y_j, satisfies the relation

$$W(y_i,y_j) = y_iy_j' - y_jy_i' = k_{ij} \exp\left[-\int P(x)\,dx\right] \tag{3}$$

where k_{ij} is a suitable constant To establish this, we begin with the hypothesis

---

† The symbol $W(y_1,y_2)$ is customarily used to denote this combination of two functions, in honor of Hoené Wronsky (1778–1853), Polish poet and mathematician who was one of the first to study determinants of this type. Such determinants are usually referred to as **Wronskians**.

that both y_i and y_j are solutions of (2) and, hence, that

$$y_i'' + P(x)y_i' + Q(x)y_i = 0$$

$$y_j'' + P(x)y_j' + Q(x)y_j = 0$$

If the first of these equations is multiplied by y_j and subtracted from y_i times the second, we obtain

$$(y_i y_j'' - y_j y_i'') + P(x)(y_i y_j' - y_j y_i') = 0 \tag{4}$$

Now,

$$\frac{dW(y_i,y_j)}{dx} = \frac{d(y_i y_j' - y_j y_i')}{dx} = (y_i y_j'' + y_i' y_j') - (y_j' y_i' + y_j y_i'')$$

$$= (y_i y_j'' - y_j y_i'')$$

Hence, Eq (4) can be written

$$\frac{dW(y_i,y_j)}{dx} + P(x)W(y_i,y_j) = 0$$

This is a very simple, separable differential equation whose solution can be written down immediately

$$W(y_i,y_j) = k_{ij} \exp\left[-\int P(x)\,dx\right]\dagger$$

where k_{ij} is an integration constant. This establishes the relation (3), which is usually known as **Abel's identity,** after the Norwegian mathematician Niels Abel (1802–1829)

Now consider the two pairs of solutions (y_3,y_1) and (y_3,y_2), where y_3 is any solution whatsoever of the homogeneous equation (2) Applying Abel's identity (3) to each of these pairs in turn, we have

$$y_3y_1' - y_1y_3' = k_{31}\exp\left[-\int P(x)\,dx\right]$$

$$y_3y_2' - y_2y_3' = k_{32}\exp\left[-\int P(x)\,dx\right]$$

In general it is possible to solve these two simultaneous equations for y_3, getting

$$y_3 = \frac{y_1k_{32}\exp\left[-\int P(x)\,dx\right] - y_2k_{31}\exp\left[-\int P(x)\,dx\right]}{y_1y_2' - y_2y_1'}$$

If we now apply Abel's identity to the denominator of the last expression, we obtain

$$y_3 = \frac{y_1k_{32}\exp\left[-\int P(x)\,dx\right] - y_2k_{31}\exp\left[-\int P(x)\,dx\right]}{k_{12}\exp\left[-\int P(x)\,dx\right]}$$

$$= \frac{k_{32}}{k_{12}}y_1 - \frac{k_{31}}{k_{12}}y_2$$

Interpreting k_{32}/k_{12} as c_1 and $-k_{31}/k_{12}$ as c_2, we have thus succeeded in exhibiting *any* solution y_3 as a linear combination $c_1y_1 + c_2y_2$ of the two particular solutions y_1 and y_2, provided only that the expression

$$y_1y_2' - y_2y_1' = W(y_1,y_2)$$

by which we had to divide in order to solve for y_3, does not vanish. Theorem 2 is thus established.

---

† Since an exponential function can never vanish, it follows that wherever $\int P(x)\,dx$ exists, the Wronskian of y_i and y_j is either never zero or identically zero, according as $k_{ij} \neq 0$ or $k_{ij} = 0$

From Theorem 2 it is clear that to find a complete solution of Eq (2) we must first find two particular solutions that have a nonvanishing Wronskian, or in other words are linearly independent (Exercise 6), and then we must form a linear combination of these solutions with arbitrary coefficients We must remember, however, that, although there are infinitely many pairs of particular solutions y_1 and y_2 which can be used as a basis for constructing a complete solution of Eq (2), neither Theorem 1 nor Theorem 2 tells us how to find them In fact there is *no* general method for solving Eq (2)* and the only procedure applicable in all cases is one which permits us to determine a second, independent solution when one solution is known

To develop this procedure, let us suppose that $y_1(x) \not\equiv 0$ is a solution of Eq (2), and let us attempt to find a function $\phi(x)$ with the property that $\phi(x)y_1(x)$ is also a solution of (2) Substituting $y = \phi(x)y_1(x)$ into Eq (2), we have

$$(y_1''\phi + 2y_1'\phi' + y_1\phi'') + P(x)(y_1'\phi + y_1\phi') + Q(x)(y_1\phi)$$

$$= [y_1'' + P(x)y_1' + Q(x)y_1]\phi + [2y_1' + P(x)y_1]\phi' + y_1\phi'' \overset{?}{=} 0$$

Now, the coefficient of ϕ in the last expression is identically zero, since, by hypothesis, y_1 is a solution of Eq (2) Hence, the last equation will be satisfied provided ϕ is chosen such that

$$y_1\phi'' + [2y_1' + P(x)y_1]\phi' = 0$$

This is a simple separable equation in ϕ', and we have

$$\frac{d\phi'}{\phi'} + \left[\frac{2y_1'}{y_1} + P(x)\right]dx = 0$$

or, integrating,

$$\ln|\phi'| + 2\ln|y_1| + \int P(x)\,dx = \ln|c|$$

Hence, combining the logarithms and taking antilogs,

$$\phi' = \frac{c\exp[-\int P(x)\,dx]}{y_1^2}$$

Integrating again, we find

$$\phi = c\int\frac{\exp[-\int P(x)\,dx]}{y_1^2}\,dx + k$$

from which we obtain, for all values of c and k, the solution

(5) $$\phi(x)y_1(x) = cy_1(x)\int\frac{\exp[-\int P(x)\,dx]}{y_1^2(x)}\,dx + ky_1(x)$$

Since this contains two arbitrary constants, it is actually a complete solution, provided that the two particular solutions from which it is constructed, namely,

---

* The nearest thing to a general solution procedure is the use of infinite series, described in Sec 9 1

$$y_1(x) \int \frac{\exp[-\int P(x)\,dx]}{y_1^2(x)}\,dx \qquad \text{and} \qquad y_1(x)$$

have a nonvanishing Wronskian. It is not difficult to show that this is always the case, although we shall leave the proof as an exercise.

EXAMPLE 1

Find a complete solution of the equation $x^2y'' + xy' - 4y = 0$, given that $y = x^2$ is one solution

Substituting the assumed solution $y = x^2\phi$ into the given differential equation, we have

$$x^2(2\phi + 4x\phi' + x^2\phi'') + x(2x\phi + x^2\phi') - 4x^2\phi = 0$$

or, simplifying,

$$x\phi'' + 5\phi' = 0$$

Separating variables, we obtain

$$\frac{d\phi'}{\phi'} + \frac{5}{x}\,dx = 0$$

Then, integrating, we get

$$\ln|\phi'| + 5\ln|x| = \ln|c| \qquad \text{or} \qquad \phi' = \frac{c}{x^5}$$

and finally, integrating again,

$$\phi = -\frac{c}{4x^4} + k$$

The complete solution is, therefore,

$$y = x^2\phi = -\frac{c}{4x^2} + kx^2$$

The solution of the nonhomogeneous equation (1) is based on the following theorem

THEOREM 3

If Y is any solution of the nonhomogeneous equation

$$y'' + P(x)y' + Q(x)y = R(x)$$

and if $c_1y_1 + c_2y_2$ is a complete solution of the homogeneous equation obtained from this by deleting the term $R(x)$, then $y = c_1y_1 + c_2y_2 + Y$ is a complete solution of the nonhomogeneous equation

PROOF Let $\bar{y}$ be any solution whatsoever of the nonhomogeneous equation (1) Then

$$\bar{y}'' + P(x)\bar{y}' + Q(x)\bar{y} = R(x)$$

and, similarly, since Y is also a solution of (1),

$$Y'' + P(x)Y' + Q(x)Y = R(x)$$

If we subtract the last two equations, we obtain

$$(\bar{y}'' - Y'') + P(x)(\bar{y}' - Y') + Q(x)(\bar{y} - Y) = 0$$

or

$$(\bar{y} - Y)'' + P(x)(\bar{y} - Y)' + Q(x)(\bar{y} - Y) = 0$$

Thus the quantity $\bar{y} - Y$ satisfies the homogeneous equation (2) and, hence, by Theorem 2, must be expressible in the form

$$\bar{y} - Y = c_1y_1 + c_2y_2$$

provided that $W(y_1,y_2) \neq 0$; that is, provided that $c_1y_1 + c_2y_2$ is a complete solution of (2), as we assumed Therefore, transposing,

$$\bar{y} = c_1y_1 + c_2y_2 + Y$$

Since $\bar{y}$ was *any* solution of the nonhomogeneous equation (1), Theorem 3 is thus established

The term Y, which can be any solution of (1) no matter how special, is called a **particular integral** of the nonhomogeneous equation The expression $c_1y_1 + c_2y_2$, which is a complete solution of the homogeneous equation corresponding to (1), is called the **complementary function** of the nonhomogeneous equation The steps to be carried out in solving an equation of the form (1) can be summarized as follows

a Delete the term $R(x)$ from the given equation, and then find two solutions of the resulting homogeneous equation which have a nonvanishing Wronskian Then combine these to form the complementary function $c_1y_1 + c_2y_2$ of the given equation.

b Find one particular solution Y of the nonhomogeneous equation itself

c Add the complementary function $c_1y_1 + c_2y_2$ found in step **a** to the particular integral Y found in step **b**, to obtain the complete solution $y = c_1y_1 + c_2y_2 + Y$ of the given equation.

In the following sections we shall investigate how these theoretical steps can be carried out when $P(x)$ and $Q(x)$ are constants; that is, when we have the so-called **linear differential equation with constant coefficients.**

EXERCISES

1 Using the one solution indicated, find a complete solution of each of the following equations

a $y'' + y = 0$ $\qquad y_1 = \sin x$
b $y'' + 3y' + 2y = 0$ $\qquad y_1 = e^{-x}$
c $(1 - 2x)y'' + 2y' + (2x - 3)y = 0$ $\qquad y_1 = e^x$
d $(2x - x^2)y'' + 2(x - 1)y' - 2y = 0$ $\qquad y_1 = x - 1$
e $x^2y'' + 4xy' - 4y = 0$ $\qquad y_1 = x$
f $x^2y'' - (x^2 + 2x)y' + (x + 2)y = 0$ $\qquad y_1 = x$

2 Verify that each of the following equations has the indicated solutions, and in each case construct two different complete solutions.

a $y'' - y = 0$ $\qquad y_1 = e^x$ $\qquad y_2 = e^{-x}$
b $y'' - 3y' + 2y = 0$ $\qquad y_1 = e^x$ $\qquad y_2 = e^{2x}$
c $y'' + y = 0$ $\qquad y_1 = \sin (x + \pi/4)$ $\qquad y_2 = \sin (x - \pi/4)$
d $y'' - 2(\cot 2x)y' = 0$ $\qquad y_1 = \sin^2 x$ $\qquad y_2 = \cos^2 x$

3 Show that the two solutions

$$y_1 \quad \text{and} \quad y_1 \int \frac{\exp\left[-\int P(x)\,dx\right]}{y_1^2}\,dx$$

of the equation $y'' + P(x)y' + Q(x)y = 0$ have a nonvanishing Wronskian

4 If the Wronskian of two functions is different from zero at every point of an interval, show that there is no point of the interval at which the functions are simultaneously zero

5 If the Wronskian of two functions is different from zero at every point of an interval, show that there is no point of the interval at which either function has a repeated zero

6 Show that, if two differentiable functions are linearly dependent, their Wronskian is equal to zero

7 Show that the converse of the assertion of Exercise 6 is false Hint Consider, for $-\infty < x < \infty$, the following pair of functions

$$y_1 = x^2 \qquad y_2 = \begin{cases} -x^2 & -\infty < x \leq 0 \\ x^2 & 0 < x < \infty \end{cases}$$

8 Show that the converse of the assertion of Exercise 6 is true over any interval on which the two functions have no common zero

9 Show that, if the Wronskian of two functions is different from zero at every point of an interval, then, between any two consecutive zeros of either of the functions in that interval there is exactly one zero of the other function [Hint Let y_1 and y_2 be the two functions, let a and b be consecutive zeros of y_1, apply Rolle's theorem to the quotient y_1/y_2 over the interval (a,b), and note the contradiction unless $y_2 = 0$ at some point between a and b]

10 Explain how Abel's identity can be used to find a second solution of the equation $y'' + P(x)y' + Q(x)y = 0$ when one solution is known Illustrate the method by applying it to parts **a** and **b** of Exercise 1

2.2

The homogeneous linear equation with constant coefficients

When $P(x)$ and $Q(x)$ are constants, the general linear second-order differential equation can be written in the standard form

$$ay'' + by' + cy = f(x) \tag{1}$$

A second standard form which is often encountered is based upon the so-called **operator notation.** In this, the symbol of differentiation d/dx is replaced by D, so that, by definition,

$$Dy \equiv \frac{dy}{dx}\dagger$$

As an immediate extension, the second derivative, which, of course, is obtained by a repetition of the process of differentiation, is written

$$D(Dy) = D^2y$$

---

† Just as the prime notation, y', y'', , may in specific instances indicate derivatives with respect to x, t, or any other independent variable, so the operator notation, Dy, D^2y, , may also indicate derivatives with respect to an independent variable other than x, depending on the context

Similarly, $\dfrac{d^3y}{dx^3} = D(D^2y) = D^3y$

$$\frac{d^4y}{dx^4} = D(D^3y) = D^4y$$

. .

Evidently, positive integral powers of D (which are the only ones we have defined) obey the usual laws of exponents

If due care is taken to see that variables are not moved across the sign of differentiation by a careless interchange of the order of factors containing variable coefficients, the operator D can be handled in many respects as though it were a simple algebraic quantity For instance, after defining $(aD^2 + bD + c)f(x)$ to mean $aD^2f(x) + bDf(x) + cf(x)$, we have, for the polynomial operator $3D^2 - 10D - 8$ and its factored equivalents,

$$(3D^2 - 10D - 8)x^2 = 3(2) - 10(2x) - 8(x^2) = 6 - 20x - 8x^2$$

$$(3D + 2)(D - 4)x^2 = (3D + 2)(2x - 4x^2)$$

$$= (6 - 24x) + (4x - 8x^2) = 6 - 20x - 8x^2$$

$$(D - 4)(3D + 2)x^2 = (D - 4)(6x + 2x^2)$$

$$= (6 + 4x) - (24x + 8x^2) = 6 - 20x - 8x^2$$

which illustrates how algebraically equivalent forms of an operator yield identical results when applied to the same function

Using the operator D, we can evidently write Eq (1) in the alternative standard form

(1a) $$(aD^2 + bD + c)y = f(x)$$

Many writers base the solution of Eq (1) upon the operational properties of the symbol D However, we shall postpone all operational methods until the chapter on the Laplace transformation, where operational calculus can be developed easily and efficiently in its proper setting

Following the theory of the last section, we first attempt to find a complete solution of the homogeneous equation

(2) $$ay'' + by' + cy = 0$$

or

(2a) $$(aD^2 + bD + c)y = 0$$

obtained from (1) or (1a) by deleting $f(x)$ In searching for particular solutions of (2), it is natural to try

$$y = e^{mx}$$

where m is a constant to be determined, because all derivatives of this function are alike except for a numerical coefficient Substituting into Eq (2) and then factoring e^{mx} from every term, we have

$$e^{mx}(am^2 + bm + c) = 0$$

as the condition to be satisfied if $y = e^{mx}$ is to be a solution. Since e^{mx} can never be zero, it is thus necessary that

$$am^2 + bm + c = 0 \tag{3}$$

This purely algebraic equation is known as the **characteristic** or **auxiliary** equation of either Eq (1) or Eq (2) In practice it is obtained not by substituting $y = e^{mx}$ into the given differential equation and then simplifying, but rather by substituting m^2 for y'', m for y', and 1 for y in the given equation, or, still more simply, by equating to zero the operational coefficient of y and then letting D play the role of m

$$aD^2 + bD + c = 0$$

The characteristic equation is a simple quadratic which will in general be satisfied by two values of m

$$m = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Using these values, say m_1 and m_2, two solutions

$$y_1 = e^{m_1x} \qquad \text{and} \qquad y_2 = e^{m_2x}$$

can be constructed. From this pair, according to Theorem 1, Sec 2 1, an infinite family of solutions

$$y = c_1y_1 + c_2y_2 = c_1e^{m_1x} + c_2e^{m_2x} \tag{4}$$

can be formed Moreover, by Theorem 2, Sec 2 1, if the Wronskian of these solutions is different from zero, then (4) is a complete solution of Eq (2), i e, it contains all possible solutions of the homogeneous equation Accordingly, we compute

$$\begin{aligned} W(y_1,y_2) = y_1y_2' - y_2y_1' &= e^{m_1x}(m_2e^{m_2x}) - e^{m_2x}(m_1e^{m_1x}) \\ &= (m_2 - m_1)e^{(m_1+m_2)x} \end{aligned}$$

Since $e^{(m_1+m_2)x}$ can never vanish, it is clear that *a complete solution of Eq (2) is always given by (4), except in the special case when $m_1 = m_2$ and the Wronskian vanishes identically.*

EXAMPLE 1

Find a complete solution of the differential equation $y'' + 7y' + 12y = 0$

The characteristic equation in this case is

$$m^2 + 7m + 12 = 0$$

and its roots are

$$m_1 = -3 \qquad \text{and} \qquad m_2 = -4$$

Since these values of m are different, a complete solution is

$$y = c_1e^{-3x} + c_2e^{-4x}$$

EXAMPLE 2

Find a complete solution of the equation $y'' + 2y' + 5y = 0$

The characteristic equation in this case is

$$m^2 + 2m + 5 = 0$$

and its roots are

$$m_1 = -1 + 2i \qquad \text{and} \qquad m_2 = -1 - 2i$$

Since these are distinct, a complete solution is

$$y = c_1e^{(-1+2i)x} + c_2e^{(-1-2i)x}$$

Although the last expression is undeniably a complete solution of the given equation, it is unsatisfactory for many practical purposes because it involves imaginary exponentials, which are awkward to handle and are not tabulated It is, therefore, a matter of considerable importance to construct a more convenient complete solution in the case in which m_1 and m_2 are conjugate complex quantities

To do this, let us suppose that

$$m_1 = p + iq \qquad \text{and} \qquad m_2 = p - iq$$

so that a complete solution as first constructed is

$$y = c_1e^{(p+iq)x} + c_2e^{(p-iq)x}$$

By factoring out e^{px}, this can be written as

$$y = e^{px}(c_1e^{iqx} + c_2e^{-iqx})$$

Now the expression in parentheses can be simplified by using the **Euler formulas** (Sec 14 7)

$$e^{i\theta} = \cos\theta + i\sin\theta$$

$$e^{-i\theta} = \cos\theta - i\sin\theta$$

taking $\theta = qx$ The result of these substitutions is

$$\begin{aligned} y &= e^{px}[c_1(\cos qx + i\sin qx) + c_2(\cos qx - i\sin qx)] \\ &= e^{px}[(c_1 + c_2)\cos qx + i(c_1 - c_2)\sin qx] \end{aligned}$$

If we now define two new arbitrary constants by the equations

$$A = c_1 + c_2 \qquad \text{and} \qquad B = i(c_1 - c_2)$$

the complete solution can finally be put in the purely real form

$$y = e^{px}(A\cos qx + B\sin qx)$$

Of course, it is not difficult to verify directly that both

$$y_1 = e^{px}\cos qx \qquad \text{and} \qquad y_2 = e^{px}\sin qx$$

are particular solutions of the homogeneous equation (2) For a completely satisfactory derivation this should now be done, since we do not yet know that our formal treatment of complex exponentials, as though they obeyed the same laws as real exponentials, is justified

EXAMPLE 2 (continued)

Applying the preceding reasoning to Example 2, it is clear that $p = -1$ and $q = 2$ Hence, the complete solution can be written

$$y = e^{-x}(A\cos 2x + B\sin 2x)$$

When the characteristic equation has equal roots, the two independent solutions normally arising from the substitution of $y = e^{mx}$ become identical, and, as we pointed out above, we do not have an adequate basis for constructing a complete solution. To find a second, independent solution in this case we use the method developed in the last section.

Let the differential equation in question be

$$y'' - 2ry' + r^2y = 0$$

so that its characteristic equation

$$m^2 - 2rm + r^2 = 0$$

has the repeated root $m_1 = r$. Then $y_1 = e^{rx}$ is one solution, and, from Eq. (5), Sec. 2.1, a second independent solution is given by

$$y_1 \int \frac{e^{-\int P(x)dx}}{y_1^2}\,dx = e^{rx} \int \frac{e^{2rx}}{(e^{rx})^2}\,dx = xe^{rx} \equiv xe^{m_1x}$$

Thus, *in the exceptional case in which the characteristic equation has equal roots, a complete solution of Eq. (2) is*

$$y = c_1e^{m_1x} + c_2xe^{m_1x}$$

EXAMPLE 3

Find a complete solution of the equation $(D^2 + 6D + 9)y = 0$.

In this case the characteristic equation

$$m^2 + 6m + 9 = 0$$

is a perfect square, with roots $m_1 = m_2 = -3$. Hence, by our last remark, a complete solution of the given equation is

$$y = c_1e^{-3x} + c_2xe^{-3x}$$

The complete process for solving the homogeneous equation (2) in all possible cases is summarized in Table 2.1.

Table 2.1

Differential equation $ay'' + by' + cy = 0$ or $(aD^2 + bD + c)y = 0$

Characteristic equation $am^2 + bm + c = 0$ or $aD^2 + bD + c = 0$

Nature of the roots of the characteristic equation	Condition on the coefficients of the characteristic equation	Complete solution of the differential equation
Real and unequal $m_1 \neq m_2$	$b^2 - 4ac > 0$	$y = c_1e^{m_1x} + c_2e^{m_2x}$
Real and equal $m_1 = m_2$	$b^2 - 4ac = 0$	$y = c_1e^{m_1x} + c_2xe^{m_1x}$
Conjugate complex $m_1 = p + iq$ $m_2 = p - iq$	$b^2 - 4ac < 0$	$y = e^{px}(A \cos qx + B \sin qx)$

In particular applications, the two arbitrary constants in the complete solution must usually be determined to fit initial conditions on y and y', or their equivalent. The following examples will clarify the procedure

EXAMPLE 4

Find the solution of the equation $y'' - 4y' + 4y = 0$ for which $y = 3$ and $y' = 4$ when $t = 0$

The characteristic equation of the differential equation is

$$m^2 - 4m + 4 = 0$$

Its roots are $m_1 = m_2 = 2$, hence, a complete solution is

$$y = c_1e^{2t} + c_2te^{2t}$$

By differentiating this, we find

$$y' = (2c_1 + c_2)e^{2t} + 2c_2te^{2t}$$

Substituting the given data into the equations for y and y', respectively, we have

$$3 = c_1 \qquad 4 = 2c_1 + c_2$$

Hence, $c_1 = 3$, $c_2 = -2$, and the required solution is

$$y = 3e^{2t} - 2te^{2t}$$

EXAMPLE 5

Find the solution of the equation $(4D^2 + 16D + 17)y = 0$ for which $y = 1$ when $t = 0$ and $y = 0$ when $t = \pi$

In this case the characteristic equation is

$$4m^2 + 16m + 17 = 0$$

and, from its roots, $m = -2 \pm \frac{1}{2}i$, we obtain the complete solution

$$y = e^{-2t}\left(A \cos \frac{t}{2} + B \sin \frac{t}{2}\right)$$

Substituting the given conditions into this equation, we find

$$1 = A \qquad \text{and} \qquad 0 = e^{-2\pi}B \qquad \text{or} \qquad B = 0$$

Hence, the required solution is

$$y = e^{-2t} \cos \frac{t}{2}$$

EXERCISES

1 What is the difference between Dy and yD?

2 Verify that

$$(D + 1)(D^2 + 2) \sin 3x = (D^2 + 2)(D + 1) \sin 3x = (D^3 + D^2 + 2D + 2) \sin 3x$$

3 Is $(D + x)(D + 2x)e^x = (D + 2x)(D + x)e^x$? Explain

4 What meaning, if any, do you think can be assigned to D^0? D^{-1}? D^{-2}?

Find a complete solution of each of the following equations

5 $y'' + y' - 2y = 0$

6 $5y'' + 6y' + y = 0$

7 $y'' - 5y = 0$

8 $y'' - 5y' = 0$

9 $(4D^2 + 4D + 1)y = 0$

10 $(9D^2 - 12D + 4)y = 0$

11 $10y'' + 6y' + y = 0$

12 $y'' + 10y' + 26y = 0$

Find the solution of each of the following equations which satisfies the given conditions

13 $y'' + 3y' - 4y = 0$ $y = 4,\ y' = -2$ when $x = 0$

14 $y'' + 4y = 0$ $y = 2,\ y' = 6$ when $x = 0$

15 $y'' - 4y = 0$ $y = 1,\ y' = -1$ when $x = 0$

16 $25y'' + 20y' + 4y = 0$ $y = y' = 0$ when $x = 0$

17 $(D^2 + 6D + 9)y = 0$ $y = 0,\ y' = 3$ when $x = 0$

18 $(D^2 + 2D + 5)y = 0$ $y = 1$ when $x = 0$, $y' = 0$ when $x = \pi$

19 $(D^2 + 2D + 5)y = 0$ $y = 1$ when $x = 0$, $y = 0$ when $x = \pi$

20 **a** Verify by direct substitution that $y_1 = e^{px} \cos qx$ and $y_2 = e^{px} \sin qx$ are solutions of the equation

$$y'' - 2py' + (p^2 + q^2)y = 0$$

b Verify that these solutions have a nonvanishing Wronskian

21 Show that, for $k \neq 0$, both $y = A \cos (kx + B)$ and $y = G \sin (kx + H)$ are complete solutions of the equation $y'' + k^2y = 0$

22 Show that, for $k \neq 0$, $y = A \cosh kx + B \sinh kx$ is a complete solution of the equation $y'' - k^2y = 0$

23 Show that $y = e^{px}(A \cosh qx + B \sinh qx)$ is a complete solution of the equation $ay'' + by' + cy = 0$, when the roots of the characteristic equation are $p \pm q$, $q \neq 0$

24 If the roots of its characteristic equation are real, show that no nontrivial solution of the equation $ay'' + by' + cy = 0$ can have more than one real zero

25 If the characteristic equation of the differential equation $ay'' + by' + cy = 0$ has distinct roots m_1 and m_2, show that

$$y = \frac{e^{m_1x} - e^{m_2x}}{m_1 - m_2}$$

is a particular solution of the equation Can we take the limit of this expression as $m_2 \to m_1$ and obtain a second solution of the differential equation when its characteristic equation has equal roots?

2.3

The nonhomogeneous equation

In the last section we learned how to solve the homogeneous equation $ay'' + by' + cy = 0$, and with this knowledge we can now obtain the complementary function of the nonhomogeneous equation

$$ay'' + by' + cy = f(x) \tag{1}$$

However, we must also have a particular integral, i e , a particular solution, of Eq (1), before we can construct its complete solution, namely,

$$y = \text{complementary function} + \text{particular integral}$$

Various procedures are available for the determination of particular solutions of Eq. (1), some applicable no matter what $f(x)$ may be, others useful only when $f(x)$ belongs to some suitably specialized class of functions It should be borne in mind, however, that, in applying Theorem 3, Sec 2 1, the important thing is not *how* we obtain a particular solution of Eq (1) but merely *that* we have one such solution Any method, from outright guessing to the most sophisticated theoretical technique, is legitimate, provided that it leads to a solution that can be checked in (1) In this section we shall introduce the so-called **method of undetermined coefficients,** which appears initially to be based on little more than guesswork, but which is readily formalized into a well-defined procedure applicable to a well-defined and very important class of cases

To illustrate the method, suppose that we wish to find a particular integral of the equation

(2) $$y'' + 4y' + 3y = 5e^{2x}$$

Since differentiating an exponential of the form e^{kx} merely reproduces the function with, at most, a change in its numerical coefficient, it is natural to "guess" that it may be possible to determine A so that

$$Y = Ae^{2x}$$

will be a solution of (2) To check this, we substitute $Y = Ae^{2x}$ into the given equation, getting

$$4Ae^{2x} + 8Ae^{2x} + 3Ae^{2x} \overset{?}{=} 5e^{2x} \qquad \text{or} \qquad 15Ae^{2x} \overset{?}{=} 5e^{2x}$$

which will be an identity if and only if $A = \frac{1}{3}$ Thus, the required particular integral is

$$Y = \tfrac{1}{3}e^{2x}$$

Now suppose that the right-hand member of (2) had been $5 \sin 2x$ Guided by our previous success we might perhaps be led to try

$$Y = A \sin 2x$$

as a particular integral Substituting this to check whether or not it can be a solution, we obtain

$$-4A \sin 2x + 8A \cos 2x + 3A \sin 2x \overset{?}{=} 5 \sin 2x$$
$$-A \sin 2x + 8A \cos 2x \overset{?}{=} 5 \sin 2x$$

and this cannot be an identity unless, simultaneously, $A = -5$ and $A = 0$, which is absurd The difficulty here, of course, is that differentiating $\sin 2x$ introduced the new function $\cos 2x$, which must also be eliminated identically from the equation resulting from the substitution of $Y = A \sin 2x$ Since the one arbitrary constant A cannot satisfy two independent conditions, it is clear that we must arrange to incorporate *two* arbitrary constants in our tentative choice for Y This is easily done by assuming

$$Y = A \sin 2x + B \cos 2x$$

which contains the necessary second parameter, yet cannot introduce any further new functions, since it already is a linear combination of *all* the independent terms that can be obtained from $\sin 2x$ by repeated differentiation The actual determination of A and B is a simple matter, for substitution into the given differential equation yields

$$(-4A \sin 2x - 4B \cos 2x) + 4(2A \cos 2x - 2B \sin 2x) + 3(A \sin 2x + B \cos 2x) = 5 \sin 2x$$

$$(-A - 8B) \sin 2x + (8A - B) \cos 2x = 5 \sin 2x$$

and for this to be an identity requires that

$$-A - 8B = 5 \qquad \text{and} \qquad 8A - B = 0$$

from which we find immediately that $A = -\frac{1}{13}$ and $B = -\frac{8}{13}$ Hence, finally,

$$Y = -\frac{\sin 2x + 8 \cos 2x}{13}$$

With these illustrations in mind we are now in a position to describe more precisely the use of the method of undetermined coefficients for finding particular integrals

rule 1 If $f(x)$ is a function for which repeated differentiation yields only a finite number of linearly independent expressions, then, in general, a particular integral Y for the nonhomogeneous equation $ay'' + by' + cy = f(x)$ can be found by

a Assuming Y to be an arbitrary linear combination of all the linearly independent terms which arise from $f(x)$ by repeated differentiation

b Substituting Y into the given differential equation

c Determining the arbitrary constants in Y in such a way that the resulting equation is identically satisfied

The class of functions $f(x)$ possessing only a finite number of linearly independent derivatives consists of the simple functions

k
x^n (n a positive integer)
e^{kx}
$\cos kx$
$\sin kx$

and any others obtainable from these by a finite number of additions, subtractions, and multiplications If $f(x)$ possesses infinitely many independent derivatives, as is the case, for instance, with the simple function $1/x$, it is occasionally convenient to assume for Y an infinite series whose terms are the respective derivatives of $f(x)$ each multiplied by an arbitrary constant However, the use of the method of undetermined coefficients in such cases involves questions of convergence which never arise when $f(x)$ has only a finite number of independent derivatives

There is one important exception to the procedure we have just been outlining, which we must now investigate Suppose, for example, that we wish to find a particular integral for the equation

$$y'' + 4y' + 3y = 5e^{-3x} \tag{3}$$

Proceeding in the way we have just described, we would start with

$$Y = Ae^{-3x}$$

getting

$$9Ae^{-3x} - 12Ae^{-3x} + 3Ae^{-3x} \overset{?}{=} 5e^{-3x}$$

$$0 \overset{?}{=} 5e^{-3x} \ (!)$$

This is obviously an impossibility, and it is important that we be able to recognize and handle such cases The source of the difficulty is easily identified. For the characteristic equation of Eq (3) is

$$m^2 + 4m + 3 = 0$$

and, since its roots are $m_1 = -3$, $m_2 = -1$, the complementary function of Eq (3) is

$$y = c_1e^{-3x} + c_2e^{-x}$$

Thus, the term on the right-hand side of (3) is proportional to a term in the complementary function; that is, it is a solution of the related homogeneous equation and, hence, can yield only 0 when it is substituted into the left member

One way in which we might attempt to avoid this difficulty would be to find a particular integral of the equation

$$y'' + 4y' + 3y = 5e^{kx} \tag{4}$$

with $k \neq -3$, and then take the limit of this solution as $k \to -3$. Thus, substituting $Y = Ae^{kx}$, as usual, we have

$$k^2Ae^{kx} + 4kAe^{kx} + 3Ae^{kx} = 5e^{kx}$$

whence,

$$A = \frac{5}{k^2 + 4k + 3} \qquad \text{and} \qquad Y = \frac{5e^{kx}}{k^2 + 4k + 3}$$

Unfortunately, the limit of this as $k \to -3$ is infinite, and so we must look further However, since Be^{-3x} is a solution of the related homogeneous equation for all values of B, it follows, taking $B = -5/(k^2 + 4k + 3)$, that

$$y_1 = \frac{-5e^{-3x}}{k^2 + 4k + 3}$$

is a particular solution of the homogeneous equation and, hence, that

$$Y - y_1 = \frac{5e^{kx} - 5e^{-3x}}{k^2 + 4k + 3}$$

is another particular integral of the nonhomogeneous equation (4). Now as $k \to -3$ [and Eq (4) approaches the given equation (3)], this function becomes an indeterminate of the form 0/0 Evaluating it by L'Hospital's rule, we find that the limit is

$$\frac{5xe^{-3x}}{-2}$$

and by direct substitution it is easily verified that this is actually a solution of Eq (3).

It is not necessary to go through this limiting process in particular cases where $f(x)$ is proportional to a term already in the complementary function, for we have the following extension of Rule 1

rule 2 If any term in the expression Y normally used to find a particular integral of the nonhomogeneous equation $ay'' + by' + cy = f(x)$ duplicates a term in the complementary function, then before it is substituted into the equation, Y must be multiplied by the lowest positive integral power of x which will eliminate all such duplications.

The results of our discussion are summarized in Table 2.2.

table 2 2

Differential equation $ay'' + by' + cy = f(x)$ or $(aD^2 + bD + c)y = f(x)$

	$f(x)$*	Necessary choice for particular integral Y†
1	α	A
2	αx^n (n a positive integer)	$A_0x^n + A_1x^{n-1} + \qquad + A_{n-1}x + A_n$
3	αe^{rx} (r either real or complex)	Ae^{rx}
4	$\alpha \cos kx$‡	$A \cos kx + B \sin kx$
5	$\alpha \sin kx$	
6	$\alpha x^n e^{rx} \cos kx$	$(A_0x^n + \qquad + A_{n-1}x + A_n)e^{rx} \cos kx$ $+ (B_0x^n + \qquad + B_{n-1}x + B_n)e^{rx} \sin kx$
7	$\alpha x^n e^{rx} \sin kx$	

---

* When $f(x)$ consists of a sum of several terms, the appropriate choice for Y is the sum of the Y expressions corresponding to these terms individually

† Whenever a term in any of the Y's listed in this column duplicates a term already in the complementary function, all terms in that Y must be multiplied by the lowest positive integral power of x sufficient to eliminate the duplication

‡ The hyperbolic functions $\cosh kx$ and $\sinh kx$ can be handled either by expressing them in terms of exponentials or by using formulas entirely analogous to those in lines 4, 5, 6, and 7

EXAMPLE 1

Find a complete solution of the equation $y'' + 9y = 2x^2 + 4x + 7$

The characteristic equation in this case is

$$m^2 + 9 = 0$$

Since its roots are $m = \pm 3i = 0 \pm 3i$, the complementary function is

$$A \cos 3x + B \sin 3x$$

According to Table 2.2, the necessary trial solutions corresponding to the respective terms in the right member of the differential equation are

$$A_0x^2 + A_1x + A_2 \qquad a_0x + a_1 \qquad \alpha_0$$

However, the last two are clearly contained in the first, and no extra generality is achieved by including them. Hence, we assume simply

$$Y = A_0x^2 + A_1x + A_2$$

Substituting this into the differential equation gives

$$2A_0 + 9(A_0x^2 + A_1x + A_2) = 2x^2 + 4x + 7$$

Equating coefficients of x^2, x, and the constant term x^0, we obtain the three equations

$$9A_0 = 2 \qquad 9A_1 = 4 \qquad 2A_0 + 9A_2 = 7$$

Hence, $\quad A_0 = \frac{2}{9} \qquad A_1 = \frac{4}{9} \qquad A_2 = \frac{59}{81}$

and so a complete solution is

$$y = A \cos 3x + B \sin 3x + \frac{18x^2 + 36x + 59}{81}$$

EXAMPLE 2

Find a complete solution of the equation $y'' + 4y' + 5y = 3e^{-2x}$

In this case the characteristic equation is

$$m^2 + 4m + 5 = 0$$

and its characteristic roots are $m_1, m_2 = -2 \pm i$. Hence, the complementary function is

$$e^{-2x}(A \cos x + B \sin x)$$

For the trial solution Y normally corresponding to the term $3e^{-2x}$, we have $Y = Ce^{-2x}$. Moreover, although each term of the complementary function contains e^{-2x} as a *factor*, e^{-2x} does not itself occur as a *term* in the complementary function, therefore, it is unnecessary and in fact incorrect to modify Y by multiplying it by any power of x. Hence, we substitute $Y = Ce^{-2x}$ into the given equation, getting

$$4Ce^{-2x} - 8Ce^{-2x} + 5Ce^{-2x} = 3e^{-2x} \qquad \text{or} \qquad Ce^{-2x} = 3e^{-2x}$$

Thus, $C = 3$, the particular integral is $Y = 3e^{-2x}$, and a complete solution is

$$y = e^{-2x}(A \cos x + B \sin x) + 3e^{-2x}$$

EXAMPLE 3

Find a complete solution of the equation $y'' + 5y' + 6y = 3e^{-2x} + e^{3x}$

The roots of the characteristic equation

$$m^2 + 5m + 6 = 0$$

are $m_1 = -2$, $m_2 = -3$. Hence, the complementary function is

$$c_1e^{-2x} + c_2e^{-3x}$$

For the trial solution Y_1 corresponding to the term $3e^{-2x}$ we would normally use Ae^{-2x}. However, e^{-2x} is already a part of the complementary function, and thus, following the second footnote to Table 2.2, we must multiply this by x before including it in Y_1. For the term e^{3x} the normal choice for a trial solution, namely, $Y_2 = Be^{3x}$, is satisfactory as it stands, since e^{3x} is not contained in the complementary function. Hence, we assume

$$Y = Y_1 + Y_2 = Axe^{-2x} + Be^{3x}$$

Substituting this into the differential equation, we have

$$(4Axe^{-2x} - 4Ae^{-2x} + 9Be^{3x}) + 5(-2Axe^{-2x} + Ae^{-2x} + 3Be^{3x}) + 6(Axe^{-2x} + Be^{3x}) = 3e^{-2x} + e^{3x}$$

or

$$Ae^{-2x} + 30Be^{3x} = 3e^{-2x} + e^{3x}$$

Equating coefficients of like terms, we find $A = 3$ and $B = \frac{1}{30}$. Hence

$$Y = 3xe^{-2x} + \frac{e^{3x}}{30}$$

and a complete solution is

$$y = c_1e^{-2x} + c_2e^{-3x} + 3xe^{-2x} + \frac{e^{3x}}{30}$$

EXAMPLE 4

Find a complete solution of the equation $y'' - 2y' + y = xe^x - e^x$.

The characteristic equation here is

$$m^2 - 2m + 1 = 0$$

and its roots are $m_1 = m_2 = 1$. Hence, the complementary function is

$$c_1e^x + c_2xe^x$$

According to line 6 of Table 2.2 (with $n = r = 1$, $k = 0$) we would normally try

$$Y = (A_0x + A_1)e^x$$

as a particular integral. (Note that the trial solution ordinarily required for $-e^x$, namely, A_1e^x, is automatically included in this choice.) However, these terms are already contained in the complementary function; hence, following the second footnote in the table, we must multiply Y by the lowest positive integral power of x which will eliminate all duplication between Y and the complementary function. This means that Y must be multiplied by x^2, since multiplying it by x would still leave the term xe^x common to Y and the complementary function. Thus we continue with the modified choice

$$Y = (A_0x^3 + A_1x^2)e^x$$

Substituting this into the differential equation, we obtain

$$[A_0x^3 + (6A_0 + A_1)x^2 + (6A_0 + 4A_1)x + 2A_1]e^x - 2[A_0x^3 + (3A_0 + A_1)x^2 + 2A_1x]e^x + (A_0x^3 + A_1x^2)e^x = xe^x - e^x$$

or

$$6A_0xe^x + 2A_1e^x = xe^x - e^x$$

This will be identically true if and only if $A_0 = \frac{1}{6}$ and $A_1 = -\frac{1}{2}$. Hence,

$$Y = \left(\frac{x^3}{6} - \frac{x^2}{2}\right)e^x$$

and so a complete solution is

$$y = c_1e^x + c_2xe^x - \frac{x^2e^x}{2} + \frac{x^3e^x}{6}$$

EXERCISES

Find a complete solution of each of the following equations:

1 $y'' + 4y' + 3y = x - 1$

2 $y'' + 3y' = \sin x + 2\cos x$

3 $y'' + y' = x + 2$

4 $y'' + y = \cos x + 3\sin 2x$

5 $y'' + 2y' + 10y = 3x^2$

6 $y'' + 4y' + 13y = \cos 3x - \sin 3x$

7 $y'' - y = e^x + 2e^{2x}$

8 $y'' + 4y' = x\sin x$

9 $(D^2 + 1)y = e^x \sin x$

10 $(D^2 + 4D + 4)y = xe^{-x}$

11 $y'' + 2y' + y = \cos^2 x$ [Hint $\cos^2 x = (1 + \cos 2x)/2$]

12 $10y'' - 6y' + y = 4\sin x\cos x$

13 a Show that $Y = -\cosh x$ is a particular integral of the equation $y'' + y' - 2y = e^{-x}$

b Determine A such that $Y = A\sinh x$ will be a particular integral of this equation

Find that solution of each of the following equations which satisfies the given conditions

14 $y'' + 4y' + 5y = 20e^x$ $\qquad y = y' = 0$ when $x = 0$

15 $y'' + 4y' + 3y = 4e^{-x}$ $\qquad y = 0,\ y' = 2$ when $x = 0$

16 $y'' + 4y' + 4y = 2x + \sin x$ $\qquad y = 1,\ y' = -1$ when $x = 0$

17 Show that $Y = (\sin\lambda t - \sin kt)/(k^2 - \lambda^2)$ is a particular integral of the equation

$$y'' + k^2y = \sin\lambda t$$

and investigate the limiting case when $\lambda \to k$

18 If y_1 and y_2 are two solutions of the nonhomogeneous equation

$$y'' + P(x)y' + Q(x)y = R(x)$$

show that $y = y_1 + y_2$ is never a solution of this equation On the other hand, if y_1 and y_2 are, respectively, solutions of the equations

$$y'' + P(x)y' + Q(x)y = R_1(x) \qquad \text{and} \qquad y'' + P(x)y' + Q(x)y = R_2(x)$$

show that $y = y_1 + y_2$ is always a solution of the equation

$$y'' + P(x)y' + Q(x)y = R_1(x) + R_2(x)$$

19 If y_1 is a solution of the homogeneous equation $y'' + P(x)y' + Q(x)y = 0$, show that the substitution of $y = \phi y_1$ will reduce the problem of finding a particular integral of the nonhomogeneous equation $y'' + P(x)y' + Q(x)y = R(x)$ to the solution of a linear equation of the first order

20 Using the method of undetermined coefficients, find a particular integral of the equation $y'' - y = 1/x$

2.4

Particular integrals by the method of variation of parameters

For certain theoretical purposes and occasionally in applications, it is desirable to be able to find a particular integral of the equation

$$ay'' + by' + cy = f(x) \tag{1}$$

in cases where the method of undetermined coefficients will not work, i e , when $f(x)$ is not one of the simple functions possessing only a finite number of independent derivatives. A procedure known as **variation of parameters** will do this for all linear equations, including the general equation with variable coefficients,

(2) $$y'' + P(x)y' + Q(x)y = R(x)$$

regardless of the form of $R(x)$, provided that a complete solution of the corresponding homogeneous equation is known This method differs from the method of undetermined coefficients in that integration rather than differentiation is involved,* which means that the price we pay for greater generality is usually the inconvenience of integrals which cannot be evaluated in terms of familiar functions

The fundamental idea behind the process is this Instead of using two arbitrary *constants* c_1 and c_2 to combine two independent solutions of the homogeneous equation

(3) $$y'' + P(x)y' + Q(x)y = 0$$

as we do in constructing the complementary function, we attempt to find two *functions* of x, say u_1 and u_2, such that

$$Y = u_1y_1 + u_2y_2$$

will be a solution of the nonhomogeneous equation (2) Having two unknown functions u_1 and u_2, we require two equations for their determination One of these will be obtained by substituting Y into the given differential equation (2), the other remains at our disposal As the analysis proceeds it will become clear what this second condition should be

From $Y = u_1y_1 + u_2y_2$ we have, by differentiation,

$$Y' = (u_1y_1' + u_1'y_1) + (u_2y_2' + u_2'y_2)$$

Another differentiation will clearly introduce second derivatives of the unknown functions u_1 and u_2, with attendant complications, unless we arrange to eliminate the first derivative terms u_1' and u_2' from Y'. This can be done if we make

(4) $$u_1'y_1 + u_2'y_2 = 0$$

which thus becomes the necessary second condition on u_1 and u_2

Proceeding now with the simplified expression

$$Y' = u_1y_1' + u_2y_2'$$

we find $$Y'' = (u_1y_1'' + u_1'y_1') + (u_2y_2'' + u_2'y_2')$$

Substituting Y, Y', and Y'' into Eq (2), we obtain

$$(u_1y_1'' + u_1'y_1' + u_2y_2'' + u_2'y_2') + P(x)(u_1y_1' + u_2y_2') + Q(x)(u_1y_1 + u_2y_2) = R(x)$$

or

$$u_1[y_1'' + P(x)y_1' + Q(x)y_1] + u_2[y_2'' + P(x)y_2' + Q(x)y_2] + u_1'y_1' + u_2'y_2' = R(x)$$

---

* This is the origin of the name *particular integral*.

The expressions in brackets vanish because, by hypothesis, both y_1 and y_2 are solutions of the homogeneous equation (3) Hence, we find for the other condition on u_1 and u_2

(5) $$u_1'y_1' + u_2'y_2' = R(x)$$

Solving Eqs (4) and (5) for u_1' and u_2', we obtain

(6) $$u_1' = -\frac{y_2}{y_1y_2' - y_2y_1'}R(x) \qquad \text{and} \qquad u_2' = \frac{y_1}{y_1y_2' - y_2y_1'}R(x)$$

The functions y_1, y_2, y_1', y_2', and $R(x)$ are all known Hence, u_1 and u_2 can be found by a single integration. With u_1 and u_2 known, the particular integral

$$Y = u_1y_1 + u_2y_2$$

is completely determined

We should notice, of course, that, if $y_1y_2' - y_2y_1' = 0$, the solution for u_1' and u_2' cannot be carried out However, $y_1y_2' - y_2y_1'$ is precisely the Wronskian of the two solutions y_1 and y_2, and, if these are independent, as we suppose them to be, then their Wronskian cannot vanish.

EXAMPLE 1

Find a complete solution of the equation $y'' + y = \sec x$

By inspection, we see that the complementary function in this case is

$$A\cos x + B\sin x$$

Hence, taking $y_1 = \cos x$ and $y_2 = \sin x$, we have from Eq (6)

$$u_1' = -\left[\frac{\sin x}{\cos x(\cos x) - \sin x(-\sin x)}\right]\sec x = -\tan x$$

$$u_2' = \left[\frac{\cos x}{\cos x(\cos x) - \sin x(-\sin x)}\right]\sec x = 1$$

Therefore, $$u_1 = -\int \tan x\,dx = \ln|\cos x| \qquad \text{and} \qquad u_2 = \int dx = x$$

and, thus, $$Y = u_1y_1 + u_2y_2 = (\ln|\cos x|)\cos x + x\sin x$$

Finally, $$y = A\cos x + B\sin x + (\ln|\cos x|)\cos x + x\sin x$$

EXERCISES

Find a complete solution of each of the following equations

1 $y'' + 4y' + 4y = e^{-2x}/x^2$

2 $4y'' + y = (x^2 - 1)/(x\sqrt{x})$

3 $y'' + 2y' + y = e^{-x}\ln|x|$

4 $y'' + 2y' + 10y = e^{-x}\sec 3x$

5 Find a particular integral of the equation $x^2y'' + xy' - y = 1/(x + 1)$, given that $y_1 = x$ and $y_2 = 1/x$ are two solutions of the related homogeneous equation

6 Find a particular integral of the equation $x^2y'' - xy' + y = 1/x$, given that $y_1 = x$ and $y_2 = x\ln|x|$ are two solutions of the related homogeneous equation

7 Find a particular integral of the equation $x^2y'' - 2xy' + 2y = x\ln|x|$, given that $y_1 = x$ and $y_2 = x^2$ are two solutions of the related homogeneous equation

8 Using the method of variation of parameters, show that the expression

$$y = A\cos kx + B\sin kx + \frac{1}{k}\int_0^x \sin k(x - s)f(s)\,ds$$

is a complete solution of the equation $y'' + k^2y = f(x)$

Find a complete solution of each of the following equations:

9 $y'' + 2ay' + (a^2 - b^2)y = f(x)$ **10** $y'' + 2ay' + (a^2 + b^2)y = f(x)$

11 $y'' + 2ay' + a^2y = f(x)$

12 Using the method of variation of parameters, find a particular integral of the equation $y'' - y = 1/x$. How does this result compare with the result of Exercise 20, Sec. 2.3?

2.5

Equations of higher order

The theory of the linear differential equation of order higher than 2,

$$y^{(n)} + P_1(x)y^{(n-1)} + \cdots + P_{n-1}(x)y' + P_n(x)y = R(x) \tag{1}$$

parallels the second-order case in all significant details. In particular, with the obvious changes required by the fact that $n > 2$, the three fundamental theorems of Sec. 2.1 hold for linear equations of all orders.* For the especially important case of the homogeneous, linear, constant-coefficient equation of order higher than 2,

$$a_0y^{(n)} + a_1y^{(n-1)} + \cdots + a_{n-1}y' + a_ny = 0 \tag{2}$$

the substitution $y = e^{mx}$ leads, as before, to the characteristic equation

$$a_0m^n + a_1m^{n-1} + \cdots + a_{n-1}m + a_n = 0 \tag{3}$$

which can be obtained in a specific problem simply by replacing each derivative by the corresponding power of m. The degree of this algebraic equation will be the same as the order of the differential equation (2); hence, counting repeated roots the appropriate number of times, we find that the number of roots $m_1, m_2, \ldots$ will equal the order of the differential equation. From these roots, the solution of the homogeneous equation can be constructed by adding together the terms that were listed in Table 2.1, Sec. 2.2, as corresponding to each of the various root types. The only extension necessary is required when the characteristic equation (3) has roots of multiplicity greater than 2. *If y_1 is the solution normally corresponding to a root m_1, and if this root occurs k (>2) times, then not only are y_1 and xy_1 solutions* (as

---

* Before Theorem 2, Sec. 2.1, can be extended to equations of higher order, it is necessary that the Wronskian of more than two functions be defined. The appropriate generalization is

$$W(y_1,y_2,\ldots,y_n) = \begin{vmatrix} y_1 & y_2 & \cdots & y_n \\ y_1' & y_2' & \cdots & y_n' \\ \cdot & \cdot & \cdot & \cdot \\ y_1^{(n-1)} & y_2^{(n-1)} & \cdots & y_n^{(n-1)} \end{vmatrix}$$

which clearly reduces to the definition of Sec. 2.1 if $n = 2$.

in the second-order case), but $x^2y_1, x^3y_1, \ldots, x^{k-1}y_1$ are also solutions and must be included in the complementary function.

For the nonhomogeneous, constant-coefficient equation

$$a_0y^{(n)} + a_1y^{(n-1)} + \cdots + a_{n-1}y' + a_ny = R(x) \tag{4}$$

it is still true that the complete solution is the sum of the complementary function, obtained by solving the associated homogeneous equation, and a particular integral. In the important case when $R(x)$ is a function possessing only a finite number of independent derivatives the particular integral can be found just as before by using the tentative choices for Y listed in Table 2.2, Sec. 2.3. Variation of parameters can be extended to those problems which the method of undetermined coefficients cannot handle. An example or two will make these ideas clear.

EXAMPLE 1

Find a complete solution of the equation $y''' + 5y'' + 9y' + 5y = 3e^{2x}$.

The characteristic equation in this case is

$$m^3 + 5m^2 + 9m + 5 = 0$$

By inspection* $m = -1$ is seen to be a root. Hence, c_1e^{-x} must be one term in the complementary function. When the factor corresponding to this root is divided out of the characteristic equation, there remains the quadratic equation

$$m^2 + 4m + 5 = 0$$

Its roots are $m = -2 \pm i$; thus, the complementary function must also contain

$$e^{-2x}(c_2 \cos x + c_3 \sin x)$$

The entire complementary function is, therefore,

$$c_1e^{-x} + e^{-2x}(c_2 \cos x + c_3 \sin x)$$

For a particular integral we try, as usual, $Y = Ae^{2x}$. Substituting this into the differential equation gives

$$(8Ae^{2x}) + 5(4Ae^{2x}) + 9(2Ae^{2x}) + 5(Ae^{2x}) = 3e^{2x} \qquad \text{or} \qquad 51Ae^{2x} = 3e^{2x}$$

Hence $$A = \frac{1}{17} \qquad Y = \frac{e^{2x}}{17}$$

and, therefore, $$y = c_1e^{-x} + e^{-2x}(c_2 \cos x + c_3 \sin x) + \frac{e^{2x}}{17}$$

EXAMPLE 2

Find a complete solution of the equation $(D^4 + 8D^2 + 16)y = -\sin x$.

The characteristic equation here is

$$m^4 + 8m^2 + 16 = 0 \qquad \text{or} \qquad (m^2 + 4)^2 = 0$$

The roots of this equation are $m = \pm 2i, \pm 2i$. Hence, the complementary function contains not only the terms

$$\cos 2x \qquad \text{and} \qquad \sin 2x$$

---

* In general, the most difficult feature of the solution of a linear, constant-coefficient differential equation of order higher than 2 is the determination of the roots of the characteristic equation. One useful procedure for doing this, *Graeffe's root-squaring process*, is discussed in the Appendix.

but also these terms multiplied by x and is, therefore,

$$c_1 \cos 2x + c_2 \sin 2x + c_3 x \cos 2x + c_4 x \sin 2x$$

To find a particular integral we try $Y = A \cos x + B \sin x$, which, on substitution into the differential equation, gives

$$(A \cos x + B \sin x) + 8(-A \cos x - B \sin x) + 16(A \cos x + B \sin x) = -\sin x$$

or

$$9A \cos x + 9B \sin x = -\sin x$$

This will be an identity if and only if $A = 0$† and $B = -\frac{1}{9}$. Therefore,

$$Y = -\frac{\sin x}{9}$$

and the complete solution is

$$y = c_1 \cos 2x + c_2 \sin 2x + c_3 x \cos 2x + c_4 x \sin 2x - \frac{\sin x}{9}$$

EXAMPLE 3

Find the solution of the equation $(D^4 + 3D^3 + 3D^2 + D)y = 2x + 8$ for which $y = y' = y'' = y''' = 0$ when $x = 0$.

The characteristic equation in this case is

$$m^4 + 3m^3 + 3m^2 + m = 0 \qquad \text{or} \qquad m(m + 1)^3 = 0$$

Its roots are $m = 0, -1, -1, -1$, hence, the complementary function, taking due account of the triple root, is

$$a + be^{-x} + cxe^{-x} + dx^2e^{-x}$$

To find a particular integral we would ordinarily assume

$$Y = Ax + B$$

However, one term in this expression (the constant B) duplicates a term already in the complementary function (the constant a). Hence, we must multiply the original choice for Y by x before using it.

Substituting the modified expression $Y = Ax^2 + Bx$ into the differential equation, we find

$$0 + 3(0) + 3(2A) + (2Ax + B) = 2x + 8$$

or

$$2Ax + (6A + B) = 2x + 8$$

For this to be identically true requires that $A = 1$ and $B = 2$. Hence, $Y = x^2 + 2x$, and the complete solution is

$$y = a + be^{-x} + cxe^{-x} + dx^2e^{-x} + x^2 + 2x \tag{5}$$

In order to impose the given initial conditions, it is necessary that we have expressions for y', y'', and y''' as well as for y. Hence we differentiate, getting

$$y' = -be^{-x} + c(e^{-x} - xe^{-x}) + d(2xe^{-x} - x^2e^{-x}) + 2x + 2 \tag{6}$$

$$y'' = be^{-x} + c(-2e^{-x} + xe^{-x}) + d(2e^{-x} - 4xe^{-x} + x^2e^{-x}) + 2 \tag{7}$$

$$y''' = -be^{-x} + c(3e^{-x} - xe^{-x}) + d(-6e^{-x} + 6xe^{-x} - x^2e^{-x}) \tag{8}$$

[†] Since the differential equation contains only derivatives of even order, we could have foreseen that Y would contain only a sine term and that $Y = B \sin x$ would be a satisfactory initial "guess."

Substituting the given conditions into Eqs (5), (6), (7), and (8), we find

$$\begin{aligned} 0 &= a + b \\ 0 &= \quad -b + c \qquad + 2 \\ 0 &= \qquad b - 2c + 2d + 2 \\ 0 &= \quad -b + 3c - 6d \end{aligned}$$

Solving these simultaneously for a, b, c, and d gives

$$a = -12 \qquad b = 12 \qquad c = 10 \qquad d = 3$$

and, finally,

$$y = -12 + 12e^{-x} + 10xe^{-x} + 3x^2e^{-x} + x^2 + 2x$$

EXERCISES

Find a complete solution of each of the following equations

1 $(D^3 + 6D^2 + 11D + 6)y = 6x - 7$

2 $(D^4 - 16)y = e^x$

3 $y''' - 2y'' - 3y' + 10y = 40 \cos x$

4 $y^{IV} + 10y'' + 9y = \cos 2x$

5 $(D^4 + 8D^2 - 9)y = x^2 + \sin 2x$

6 $(D^3 + D^2 + 3D - 5)y = e^x$

7 $(D^4 - D)y = x^2$

8 $(D^6 - 64)y = 16 \sin 2x$

Find that solution of each of the following equations which satisfies the given conditions

9 $(D^3 + 2D^2 - D - 2)y = \sin x \qquad y = y' = y'' = 0$ when $x = 0$

10 $(D^4 - 2D^3 + 2D^2 - 2D + 1)y = e^{-x} \qquad y = y' = y'' = y''' = 0$ when $x = 0$

11 $(D^3 - 2D^2 + D - 2)y = x \qquad y = y' = y'' = 1$ when $x = 0$

12 For what values of λ, if any, does the equation $y^{IV} - \lambda^4 y = 0$ have a nontrivial solution satisfying the conditions $y = y' = 0$ when $x = 0$ and $y'' = y''' = 0$ when $x = 1$? (Hint The work is easier if a complete solution containing trigonometric and hyperbolic functions is used instead of one containing trigonometric and exponential functions)

13 Using the method of variation of parameters, obtain a formula for a particular integral of the equation

$$(D^3 - 6D^2 + 11D - 6)y = f(x)$$

14 If three functions are linearly dependent, prove that their Wronskian is identically zero

15 Prove that the Wronskian of the functions $e^{m_1 x}$, $e^{m_2 x}$, and $e^{m_3 x}$ is different from zero if and only if m_1, m_2, and m_3 are all different

2.6

Applications

Linear differential equations with constant coefficients find their most important application in the study of electrical circuits and vibrating mechanical systems. So useful to engineers are the results of this analysis that we shall devote an entire chapter to its major features. However, there are also other applications of considerable interest, and, although we cannot discuss them at length, we shall conclude this chapter with a few typical examples

One important field in which linear differential equations often arise is the study of the bending of beams. When a beam is bent it is obvious that the fibers near the concave surface of the

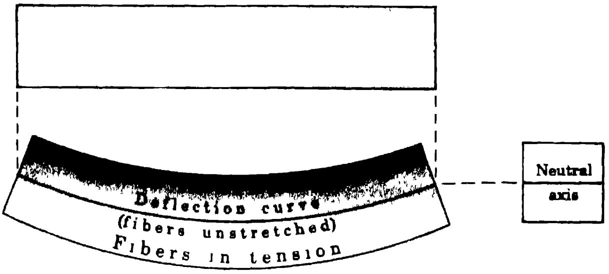


FIGURE 2.1
A beam before and after bending

beam are compressed whereas those near the convex surface are stretched. Somewhere between these regions of compression and tension there must, from considerations of continuity, be a surface of fibers which are neither compressed nor stretched. This is known as the **neutral surface** of the beam, and the curve of any particular fiber in this surface is known as the **elastic curve** or **deflection curve** of the beam. The line in which the neutral surface is cut by any plane cross section of the beam is known as the **neutral axis** of that cross section (Fig. 2.1).

The loads which cause a beam to bend may be of two sorts. They may be concentrated at one or more points along the beam, or they may be continuously distributed with a density $w(x)$ known as the **load per unit length.** In either case we have two important related quantities. One is the **shear** $V(x)$ at any point along the beam, which is defined as the algebraic sum of all the transverse forces which act on the beam on the positive side of the point in question (Fig. 2.2). The other is the **moment** $M(x)$, which is defined as the total moment produced at a general point along the beam by all the forces, transverse or not, which act on the beam on one side or the other of the point in question. We shall consider the load per unit length and the shear to be positive if they act in the direction of the negative y-axis. The moment we shall take to be positive if it acts to bend the beam so that it is concave toward the positive y-axis. With these conventions of sign (which are not universally adopted) it is shown in the study of strength of materials that the deflection of the beam $y(x)$

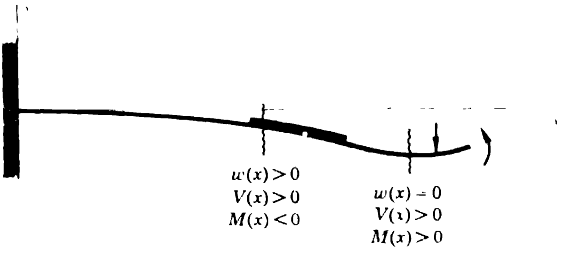


FIGURE 2.2
Plot showing the conventions for the signs of the moment, shear, and load per unit length at a general point of a beam

satisfies the second-order differential equation

$$EIy'' = M \tag{1}$$

where E is the modulus of elasticity of the material of the beam, and I, which may be a function of x, is the moment of inertia of the cross-section area of the beam about the neutral axis If the beam bears only transverse loads, it can be shown further that we have the two additional relations

$$\frac{dM}{dx} = \frac{d(EIy'')}{dx} = V \tag{2}$$

$$\frac{d^2M}{dx^2} = \frac{dV}{dx} = \frac{d^2(EIy'')}{dx^2} = -w \tag{3}$$

In most elementary applications the moment M is an explicit function of x, hence, Eq (1) can be solved and the deflection $y(x)$ determined simply by performing two integrations However, in problems in which the load has a component in the direction of the length of the beam, M depends on y, and Eq (1) can be solved only through the use of techniques from the field of differential equations An interesting example of this sort is provided by the classic problem of the buckling of a slender column

EXAMPLE 1

A long, slender column of length L and uniform cross section whose ends are constrained to remain in the same vertical line but are otherwise free (i e , are able to turn) is compressed by a load F Determine the possible deflection curves of the column and the loads required to produce each one

Let coordinates be chosen as shown in Fig 2 3 Then, clearly, the moment arm of the load F about a general point P on the deflection curve of the beam is y, hence, Eq (1) becomes

$$EIy'' = -Fy \tag{4}$$

the minus sign indicating that, when y is positive (as shown), the moment is negative, since it has produced a deflection curve which is convex toward the positive y-axis

By hypothesis, the column is of uniform cross section, hence, the moment of inertia I is a constant Therefore, (4) is a constant-coefficient differential equation and can be solved by the

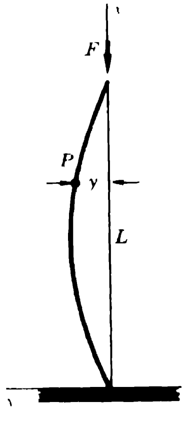


FIGURE 2 3 A slender column buckling under a vertical load methods of Sec. 2 2 Accordingly, we set up the characteristic equation

$$EIm^2 + F = 0$$

and solve it, getting

$$m = \pm \sqrt{\frac{F}{EI}}\, i$$

Hence, a complete solution of (4) is

$$y = A \cos \sqrt{\frac{F}{EI}}\, x + B \sin \sqrt{\frac{F}{EI}}\, x \tag{5}$$

To determine the constants A and B, we have the information that $y = 0$ when $x = 0$ and also when $x = L$ Substituting the first of these into Eq (5), we see at once that $A = 0$ Substituting the second, we obtain the equation

$$0 = B \sin \sqrt{\frac{F}{EI}}\, L$$

Since $\sin \sqrt{F/EI}\, L$ is in general not equal to zero, it follows that $B = 0$, which, since we have already found $A = 0$, means that $y \equiv 0$ However, if the load F has just the right value to make $\sqrt{F/EI}\, L = n\pi$, then the last equation will be satisfied without B being 0, and equilibrium is then possible in a deflected position defined by

$$y = B \sin \frac{n\pi x}{L}$$

Since n can take on any of the values 1, 2, 3, , there are thus infinitely many different **critical loads**

$$F_n = \left(\frac{n\pi}{L}\right)^2 EI$$

each with its own particular deflection curve For values of F below the lowest critical load, the column will remain in its undeflected vertical position or, if displaced slightly from it, will return to it as an equilibrium configuration For values of F above the lowest critical load and different from the higher critical loads, the column can theoretically remain in a vertical position, but the equilibrium is unstable, and, if the column is deflected slightly, it will not return to a vertical position but will continue to deflect until it collapses Thus only the lowest critical load is of much practical significance

In many physical systems vibratory motion is possible but undesirable In such cases it is important to know the frequency at which vibration *could* take place in order that periodic external influences that might be in resonance with the natural frequency of the system can be avoided For simple linear systems in which (as is usually the case) friction is neglected, the underlying differential equation is eventually reducible to the form

$$y'' + \omega^2 y = 0$$

Since the complete solution of this equation is

$$y = A \cos \omega t + B \sin \omega t$$

and since both $\cos \omega t$ and $\sin \omega t$ represent periodic behavior of frequency

$$\omega \text{ rad/unit time} \qquad \text{or} \qquad \frac{\omega}{2\pi} \text{ cycles/unit time}$$

it is clear that the frequency can be read just as well from the differential equation itself as from any of its solutions, general or particular. The important part of such a frequency calculation, then, is the formulation of the differential equation and not its solution

EXAMPLE 2

A weight W_2 is suspended from a pulley of weight W_1, as shown in Fig 2 4 Constraints, which need not be specified, prevent any swinging of the system and permit it to move only in the vertical direction If a spring of **modulus** k, that is, a spring requiring k units of force to stretch it one unit of length, is inserted in the otherwise inextensible cable which supports the pulley, find the frequency with which the system will vibrate in the vertical direction if it is displaced slightly from its equilibrium position Friction between the cable and the pulley prevents any slippage, but all other frictional effects are to be neglected

As coordinate to describe the system we choose the vertical displacement y of the center of the pulley, the downward direction being taken as positive Now when the center of the pulley moves a distance y, the length of the spring must change by $2y$ Moreover, as this happens, the pulley must rotate through an angle

$$\theta = \frac{y}{R} \qquad \text{and} \qquad \frac{d\theta}{dt} = \frac{1}{R}\frac{dy}{dt}$$

It will be convenient to formulate the differential equation governing this problem through the use of the so-called **energy method** From the fundamental law of the conservation of energy, it follows that *if no energy is lost through friction or other irreversible changes, then in a mechanical system the sum of the instantaneous potential and kinetic energies must remain constant* In the present problem the potential energy consists of two parts (a) the potential energy of the weights W_1 and W_2 due to their position in the gravitational field and (b) the potential energy stored in the stretched spring Taking the equilibrium position of the system as the reference level for potential energy, we have for (a)

$$(\text{PE})_a = -(W_1 + W_2)y \tag{6}$$

the minus sign indicating that a positive y corresponds to a lowering of the weights and, hence, a decrease in the potential energy The potential energy stored in the spring is simply the amount of work required to stretch the spring from its equilibrium elongation, say δ, to its instantaneous

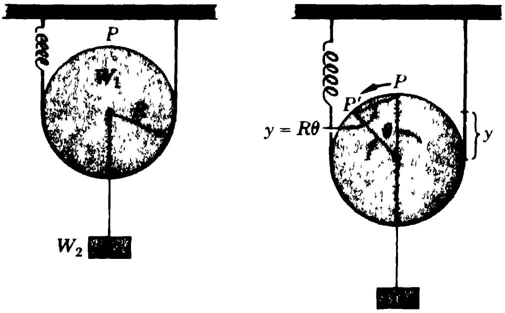


FIGURE 2 4
An unusual spring-suspended weight in equilibrium and after a vertical displacement

elongation $\delta + 2y$ Since the force in the spring at any time is

$$F = \text{elongation} \times \text{force per unit elongation} = sk$$

we have for the potential energy of type b

$$(7) \qquad (\mathrm{PE})_b = \int_{s_1}^{s_2} F\,ds = \int_{\delta}^{\delta+2y} ks\,ds = k\frac{s^2}{2}\Big|_{\delta}^{\delta+2y} = 2ky^2 + 2k\delta y$$

The kinetic energy likewise consists of two parts (a) the energy of translation of the weights W_1 and W_2, namely,

$$(8) \qquad (\mathrm{KE})_a = \frac{1}{2}\frac{W_1 + W_2}{g}(\dot{y})^2\dagger$$

and (b) the energy of rotation of the pulley, namely,

$$(9) \qquad (\mathrm{KE})_b = \frac{1}{2}I(\dot{\theta})^2 = \frac{1}{2}\frac{W_1}{g}\cdot\frac{R^2}{2}\left(\frac{\dot{y}}{R}\right)^2 = \frac{W_1}{4g}(\dot{y})^2$$

The conservation of energy now requires that

$$\text{Kinetic energy} + \text{potential energy} = \text{constant}$$

or, substituting from Eqs (6), (7), (8), and (9),

$$\frac{W_1}{4g}(\dot{y})^2 + \frac{W_1 + W_2}{2g}(\dot{y})^2 + (2ky^2 + 2k\delta y) - (W_1 + W_2)y = C$$

Differentiating this with respect to time, we have

$$\frac{W_1}{2g}\dot{y}\ddot{y} + \frac{W_1 + W_2}{g}\dot{y}\ddot{y} + 4ky\dot{y} + 2k\delta\dot{y} - (W_1 + W_2)\dot{y} = 0$$

or, dividing out $\dot{y}$ (which surely cannot be identically zero when the system is in motion) and collecting terms,

$$\frac{3W_1 + 2W_2}{2g}\ddot{y} + 4ky = (W_1 + W_2) - 2k\delta = 0$$

the terms on the right equaling zero since the elongation δ of the spring in its equilibrium position is

$$\delta = \frac{W_1 + W_2}{2k}$$

The differential equation describing the vertical movement of the system is, therefore,

$$\ddot{y} + \frac{8kg}{3W_1 + 2W_2}y = 0$$

From this, as we pointed out above, we can immediately read the natural frequency of the system, namely,

$$\frac{1}{2\pi}\sqrt{\frac{8kg}{3W_1 + 2W_2}} \quad \text{cycles/unit time}$$

In general, differential equations with variable coefficients are very difficult to solve and rarely can be solved in terms of elementary functions However, there is one important linear differential equation with variable coefficients which can always be reduced by a suitable substitution to a linear equation with

---

† In problems in dynamics, first and second derivatives *with respect to time* are often indicated by placing one and two dots, respectively, over the variable in question

constant coefficients and hence solved without difficulty This is the so-called **equation of Euler***

(10) $$a_0x^ny^{(n)} + a_1x^{n-1}y^{(n-1)} + \cdots + a_{n-1}xy' + a_ny = 0$$

in which the coefficient of each derivative is proportional to the corresponding power of the independent variable If we change the independent variable from x to z by means of the substitution

$$x = e^z \qquad \text{or} \qquad z = \ln x$$

Eq (10) becomes an equation in y and z with constant coefficients which can then be solved by the methods of Sec 2.5 Finally, replacing z by $\ln x$ in the solution of the transformed equation we obtain the solution of the original differential equation

EXAMPLE 3

Find a complete solution of the differential equation

$$x^3\frac{d^3y}{dx^3} + 4x^2\frac{d^2y}{dx^2} - 5x\frac{dy}{dx} - 15y = 0$$

Under the transformation $x = e^z$ or $z = \ln x$ we have

$$\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx} = \frac{1}{x}\frac{dy}{dz}$$

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{1}{x}\frac{dy}{dz}\right) = -\frac{1}{x^2}\frac{dy}{dz} + \frac{1}{x}\frac{d^2y}{dz^2}\frac{dz}{dx} = -\frac{1}{x^2}\frac{dy}{dz} + \frac{1}{x^2}\frac{d^2y}{dz^2}$$

$$\frac{d^3y}{dx^3} = \frac{d}{dx}\left[\frac{1}{x^2}\left(-\frac{dy}{dz} + \frac{d^2y}{dz^2}\right)\right] = -\frac{2}{x^3}\left(-\frac{dy}{dz} + \frac{d^2y}{dz^2}\right) + \frac{1}{x^2}\left(-\frac{d^2y}{dz^2} + \frac{d^3y}{dz^3}\right)\frac{dz}{dx}$$

$$= \frac{2}{x^3}\frac{dy}{dz} - \frac{3}{x^3}\frac{d^2y}{dz^2} + \frac{1}{x^3}\frac{d^3y}{dz^3}$$

Substituting these into the given differential equation, we have

$$x^3\left[\frac{1}{x^3}\left(2\frac{dy}{dz} - 3\frac{d^2y}{dz^2} + \frac{d^3y}{dz^3}\right)\right] + 4x^2\left[\frac{1}{x^2}\left(-\frac{dy}{dz} + \frac{d^2y}{dz^2}\right)\right] - 5x\left(\frac{1}{x}\frac{dy}{dz}\right) - 15y = 0$$

or, simplifying and collecting terms,

$$\frac{d^3y}{dz^3} + \frac{d^2y}{dz^2} - 7\frac{dy}{dz} - 15y = 0$$

The characteristic equation of the last equation is

$$m^3 + m^2 - 7m - 15 \equiv (m - 3)(m^2 + 4m + 5) = 0$$

From its roots, $m_1 = 3$, $m_2 = -2 + i$, $m_3 = -2 - i$, we obtain the complete solution

$$y = c_1e^{3z} + e^{-2z}(c_2\cos z + c_3\sin z)$$

Finally, replacing z by $\ln x$, we have

$$y = c_1e^{3\ln x} + e^{-2\ln x}[c_2\cos(\ln x) + c_3\sin(\ln x)]$$

$$= c_1x^3 + \frac{1}{x^2}[c_2\cos(\ln x) + c_3\sin(\ln x)]$$

---

* Also called **Cauchy's equation**, after the French mathematician Augustin Louis Cauchy (1789–1857).

EXERCISES

Find a complete solution of each of the following equations

1 $x^3y''' + 2x^2y'' - xy' + y = 0$ 2 $x^3y''' - 3x^2y'' + 7xy' - 8y = 0$

3 $2x^2\dfrac{d^2y}{dx^2} + 5x\dfrac{dy}{dx} + y = 3x + 2$

4 Since e^z is always positive, does the use of the substitution $x = e^z$ mean that an Euler equation can be solved only for positive values of x? How can a solution be obtained which will be valid for negative values of x?

5 If $x = e^z$ and if $\dfrac{d}{dz} \equiv D$, establish the operational equivalences

$$x\frac{d}{dx} = D$$

$$x^2\frac{d^2}{dx^2} = D(D-1)$$

$$x^3\frac{d^3}{dx^3} = D(D-1)(D-2)$$

$$\cdot \qquad \cdot$$

Explain how these formulas can be used to shorten the work of solving an Euler equation

6 a Show that the substitution $Ax + B = e^z$, or $z = \ln(Ax + B)$, will reduce the equation

$$a(Ax + B)^2\frac{d^2y}{dx^2} + b(Ax + B)\frac{dy}{dx} + y = 0$$

to a linear equation with constant coefficients Do you think that, for $n > 2$, the equation

$$a_0(Ax + B)^n\frac{d^ny}{dx^n} + a_1(Ax + B)^{n-1}\frac{d^{n-1}y}{dx^{n-1}} + \qquad + a_{n-1}(Ax + B)\frac{dy}{dx} + a_ny = 0$$

can be solved in a similar fashion?

b Find a complete solution of the equation

$$(x - 2)^2\frac{d^2y}{dx^2} + 2(x - 2)\frac{dy}{dx} - 6y = 0$$

7 A circular cylinder of radius r and height h, made of material weighing w lb/in ^{3}, floats in water in such a way that its axis is always vertical Neglecting all forces except gravity and the buoyant force of the water, as given by the principle of Archimedes, determine the period with which the cylinder will vibrate in the vertical direction if it is depressed slightly from its equilibrium position and released

8 A cylinder weighing 50 lb floats in water with its axis vertical When depressed slightly and released, it vibrates with period 2 sec Neglecting all frictional effects, find the diameter of the cylinder

9 A straight hollow tube rotates about its mid-point with constant angular velocity ω, the rotation taking place in a horizontal plane A pellet of mass m slides without friction in the interior of the tube Find the equation of the radial motion of the pellet until it emerges from the tube, assuming that it starts from rest at a radial distance a from the mid-point of the tube

10 A straight hollow tube rotates about its mid-point with constant angular velocity ω, the rotation taking place in a vertical plane Show that if the initial conditions are properly chosen, a pellet sliding without friction in the tube will never be ejected but will execute simple harmonic motion within the tube

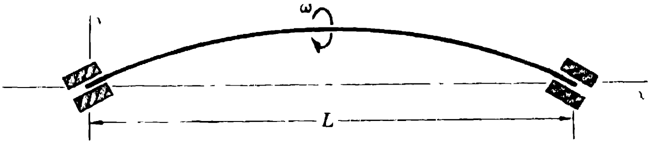
11 Neglecting the effect of its own weight, show that the deflection of a uniform cantilever beam at the point $x = x_0$ due to a unit load at the point $x = x_1$ is equal to the deflection at $x = x_1$ due to a unit load at $x = x_0$

12 A uniform cantilever beam of length L is subjected to an oblique tensile force F at the free end Find the tip deflection as a function of the angle θ between the direction of the force and the initial direction of the beam

13 A long, slender column of uniform cross section is built in rigidly at its base Its upper end, which is free to move out of line, bears a vertical load F Determine the possible deflection curves and the load required to produce each one

14 A uniform shaft of length L rotates about its axis with constant angular velocity ω The ends of the shaft are held in bearings which are free to swing out of line, as shown in Fig 2 5,

FIGURE 2 5



if the shaft deflects from its neutral position Show that there are infinitely many critical speeds at which the shaft can rotate in a deflected position, and find these speeds and the associated deflection curves [Hint During rotation, centrifugal force applies a load per unit length given by

$$w(x) = -\frac{\rho A \omega^2}{g} y$$

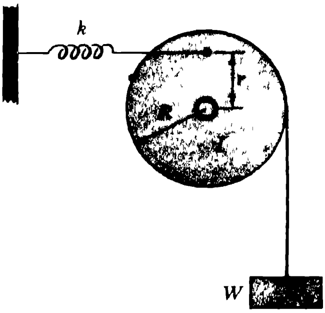
where A is the cross-section area of the shaft and ρ is the density of the material of the shaft Substitute this into Eq (3), solve the resulting differential equation, and then impose the conditions that at $x = 0$ and at $x = L$ the deflection of the shaft and the moment are zero]

15 Work Exercise 14 if the bearings are fixed in position and cannot swing out of line

16 A cantilever beam has the shape of a solid of revolution whose radius varies as $\sqrt{x}$, where x is the distance from the free end of the beam A tensile force F is applied at the free end of the beam at an angle of 45° with the initial direction of the beam Find the deflection curve of the beam

17 A weight W hangs by an inextensible cord from the circumference of a pulley of radius R and moment of inertia I The pulley is prevented from rotating freely by a spring of modulus k, attached as shown in Fig 2 6 Considering only displacements so small that the departure of the spring from the horizontal can be neglected, and neglecting all friction, determine the natural frequency of the oscillations that occur when the system is slightly disturbed (Hint Use the energy method to obtain the differential equation of the system)

FIGURE 2 6



18 Under the assumption of very small motions and neglecting friction, determine the natural frequency of the system shown in Fig. 2.7 if the bar is of uniform cross section, absolutely rigid, and of weight w.

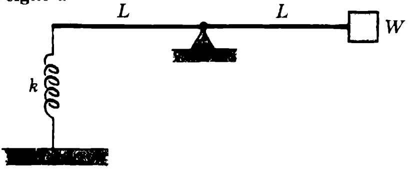


FIGURE 2.7

19 Under the assumption of very small motions and neglecting friction, determine the natural frequency of the system shown in Fig. 2.8 if the bar is of uniform cross section, absolutely rigid, and of weight w.

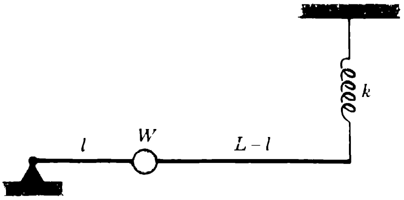


FIGURE 2.8

20 A perfectly flexible cable of length $2L$, weighing w lb/ft, hangs over a frictionless peg of negligible diameter. At $t = 0$ the cable is released from rest in a position in which the portion hanging on one side is a ft longer than that on the other. Find the equation of motion of the cable as it slips over the peg.

21 A perfectly flexible cable of length L and weighing w lb/ft lies in a straight line on a frictionless table top, a ft of the cable hanging over the edge. At $t = 0$ the cable is released and begins to slide off the edge of the table. Assuming that the height of the table is greater than L, determine the motion of the cable until it leaves the table top.

22 A perfectly flexible cable of length L, weighing w lb/ft, hangs over a pulley as shown in Fig. 2.9. The radius of the pulley is R, and its moment of inertia is I. Friction between the cable and the pulley prevents any relative slipping, although the pulley is free to turn without appreciable friction. At $t = 0$ the cable is released from rest in a position in which the portion hanging on one side is a ft longer than that hanging on the other. Determine the motion of the cable until the short end first makes contact with the pulley.

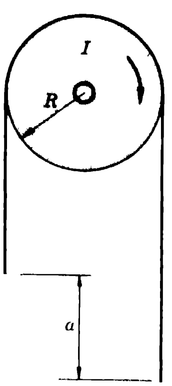
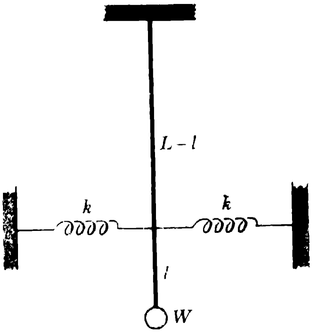


FIGURE 2.9

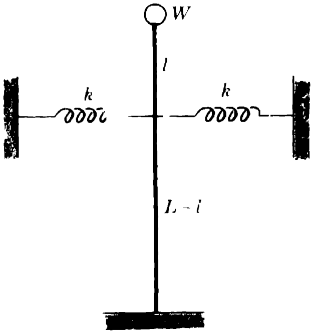
23 Neglecting friction and assuming angular displacements θ so small that θ is a satisfactory approximation to $\sin\theta$ and $\theta^2/2$ is a satisfactory approximation to $1 - \cos\theta$, find the natural frequency of the system shown in Fig 2 10, if the bar is of uniform cross section, absolutely rigid, and of weight w

FIGURE 2 10



24 Neglecting friction and assuming angular displacements θ so small that θ is a satisfactory approximation to $\sin\theta$ and $\theta^2/2$ is a satisfactory approximation to $1 - \cos\theta$, find the natural frequency of the system shown in Fig 2 11, if the bar is of uniform cross section,

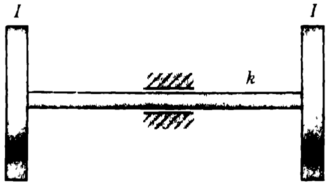
FIGURE 2 11



absolutely rigid, and of weight w In what significant respect does this system differ from that discussed in Exercise 23?

25 a Two disks each of moment of inertia I are connected by an elastic shaft of modulus k, that is, a shaft which requires k units of torque to twist one end through an angle of one radian with respect to the other end The system is mounted in a frictionless bearing, as shown in Fig 2 12 Neglecting the moment of inertia of the shaft, find the natural frequency

FIGURE 2 12



with which the disks will oscillate if they are twisted through equal but opposite angles and then released

b What is the natural frequency of the system if the moments of inertia of the disks are respectively I_1 and I_2? (Hint Not only does the total energy of the system remain constant, but so does the total angular momentum)

CHAPTER THREE

Simultaneous Linear Differential Equations

3.1

Introduction

In many problems in applied mathematics there are not one but several dependent variables, each a function of a single independent variable, usually time The formulation of such a problem in mathematical terms frequently leads to a system of simultaneous linear differential equations, as many equations as there are dependent variables

There are various methods of solving such systems In one, which bears a strong resemblance to the solution of systems of simultaneous algebraic equations, the system is reduced by successive elimination of the unknowns until a single differential equation remains This is solved, and then, working backward, the solutions for the other variables are found, one by one, until the problem is completed A second method, which amounts to considering the system as a single matric differential equation, generalizes the ideas of complementary function and particular integral and, through their use, obtains solutions for all the variables at the same time Finally, the use of the Laplace transformation provides a straightforward operational procedure for solving systems of linear differential equations with constant coefficients, which is probably preferable in most applications to either of the other methods

In this chapter we shall attempt through examples to present the first two methods, leaving the third to Chap 7, where we shall discuss the Laplace transformation and its applications in detail

3.2

The reduction of a system to a single equation

Consider the following system of equations.

$$2\frac{dx}{dt} + x + 3\frac{dy}{dt} + y = e^{-t}$$

$$\frac{dx}{dt} + 5x + \frac{dy}{dt} + 7y = t$$

If we subtract twice the second equation from the first, we obtain

(1) $$-9x + \frac{dy}{dt} - 13y = e^{-t} - 2t$$

If we subtract the second equation from 5 times the first, we obtain

(2) $$9\frac{dx}{dt} + 14\frac{dy}{dt} - 2y = 5e^{-t} - t$$

Finally, if we differentiate Eq (1) and add it to Eq (2), all occurrences of x will be eliminated and we shall have an equation in y alone

$$\frac{d^2y}{dt^2} + \frac{dy}{dt} - \quad = 4e^{-t} - t - 2$$

It is now a simple matter to solve this equation by the methods of Chap 2, and we find without difficulty

(3) $$y = c_1e^t + c_2e^{-2t} + \frac{t}{2} + \frac{5}{4} - 2e^{-t}$$

Various possibilities are available for finding x By far the simplest is to use Eq (1), which gives x directly in terms of y and its derivative Thus,

$$\begin{aligned} x &= \frac{1}{9}\left(\frac{dy}{dt} - 13y + 2t - e^{-t}\right) \\ &= \frac{1}{9}\Bigg[\left(c_1e^t - 2c_2e^{-2t} + \frac{1}{2} + 2e^{-t}\right) \\ &\qquad\qquad - 13\left(c_1e^t + c_2e^{-2t} + \frac{t}{2} + \frac{5}{4} - 2e^{-t}\right) + 2t - e^{-t}\Bigg] \end{aligned}$$

(4) $$= -\frac{4}{3}c_1e^t - \frac{5}{3}c_2e^{-2t} - \frac{t}{2} - \frac{7}{4} + 3e^{-t}$$

Equations (3) and (4) constitute a complete solution of the original system

In general, the steps in the reduction of a system of equations to a single equation are not so obvious as they were in the example we have just worked For this reason it is frequently convenient to rewrite the given equations in the D notation Then, if we regard the operational coefficients of the variables as ordinary algebraic coefficients, the method of elimination will usually be apparent Still more systematically, determinants can be used to obtain the single equation satisfied by any one of the unknowns, very much as in the case of linear algebraic equations

Suppose, for definiteness, that we have the second-order system

$$(a_{11}D^2 + b_{11}D + c_{11})x + (a_{12}D^2 + b_{12}D + c_{12})y = \phi_1(t)$$
$$(a_{21}D^2 + b_{21}D + c_{21})x + (a_{22}D^2 + b_{22}D + c_{22})y = \phi_2(t)$$

or, more compactly,

$$P_{11}(D)x + P_{12}(D)y = \phi_1(t)$$
$$P_{21}(D)x + P_{22}(D)y = \phi_2(t)$$

where the P's denote the polynomial operators which act on x and y. If these were, as indeed they appear to be, two algebraic equations in x and y, we could eliminate y at once by subtracting $P_{12}(D)$ times the second equation from $P_{22}(D)$ times the first equation, getting

(5) $$[P_{11}(D)P_{22}(D) - P_{12}(D)P_{21}(D)]x = P_{22}(D)\phi_1(t) - P_{12}(D)\phi_2(t)$$

Moreover, this procedure is clearly valid even though the system consists of differential equations rather than algebraic equations For "multiplying" the first equation by

$$P_{22}(D) \equiv a_{22}D^2 + b_{22}D + c_{22}$$

is simply a way of performing in one step the operations of adding a_{22} times the second derivative of the equation and b_{22} times the first derivative of the equation to c_{22} times the equation itself, and these steps are individually well defined and completely correct Similarly, "multiplying" the second equation by

$$P_{12}(D) \equiv a_{12}D^2 + b_{12}D + c_{12}$$

merely furnishes in one step the sum of a_{12} times the second derivative of the equation, b_{12} times the first derivative of the equation, and c_{12} times the equation itself Finally, the subtraction of the two equations obtained by the "multiplications" we have just described eliminates y and each of its derivatives because these operations produce in each equation exactly the same combination of y and its various derivatives Similarly, of course, x can be eliminated from the system by subtracting $P_{21}(D)$ times the first equation from $P_{11}(D)$ times the second, leaving a differential equation from which y can be found at once

The preceding observations can easily be formulated in determinant notation In fact, the (operational) coefficient of x in Eq (5) is simply the determinant of the (operational) coefficients of the unknowns in the original system, namely,

$$\begin{vmatrix} P_{11}(D) & P_{12}(D) \\ P_{21}(D) & P_{22}(D) \end{vmatrix}$$

Furthermore, the right-hand side of (5) can be identified as the expanded form of the determinant

$$\begin{vmatrix} \phi_1(t) & P_{12}(D) \\ \phi_2(t) & P_{22}(D) \end{vmatrix}$$

provided we keep in mind that the operators $P_{12}(D)$ and $P_{22}(D)$ must operate on $\phi_2(t)$ and $\phi_1(t)$, respectively, and hence the diagonal products must be interpreted to mean

$$P_{22}(D)\phi_1(t) \qquad \text{and} \qquad P_{12}(D)\phi_2(t)$$

and not $\quad \phi_1(t)P_{22}(D) \quad$ and $\quad \phi_2(t)P_{12}(D)$

Thus, Eq (5) can be written in the form

$$\begin{vmatrix} P_{11}(D) & P_{12}(D) \\ P_{21}(D) & P_{22}(D) \end{vmatrix} x = \begin{vmatrix} \phi_1(t) & P_{12}(D) \\ \phi_2(t) & P_{22}(D) \end{vmatrix} \tag{6}$$

which is precisely what Cramer's rule (Theorem 7, Sec 10 5) would yield if applied to the given system as though it were purely algebraic In just the same way, the result of eliminating x from the original system, namely,

$$[P_{11}(D)P_{22}(D) - P_{12}(D)P_{21}(D)]y = P_{11}(D)\phi_2(t) - P_{21}(D)\phi_1(t)$$

can be written

$$\begin{vmatrix} P_{11}(D) & P_{12}(D) \\ P_{21}(D) & P_{22}(D) \end{vmatrix} y = \begin{vmatrix} P_{11}(D) & \phi_1(t) \\ P_{21}(D) & \phi_2(t) \end{vmatrix} \tag{7}$$

The use of Cramer's rule to obtain the differential equations satisfied by the individual dependent variables is in no way restricted to the case of two equations in two unknowns Exactly the same procedure can be applied to systems of any number of equations, regardless of the degrees of the polynomial operators which appear as the coefficients of the unknowns Moreover, as Eqs (6) and (7) illustrate, the polynomial operators appearing in the left members of the equations which result when the original system is "solved" for the various unknowns are identical Hence the characteristic equations of these differential equations are identical, and, therefore, except for the presence of different arbitrary constants, the complementary functions in the solutions for the various unknowns are all the same The constants in these complementary functions are not all independent, however, and relations will always exist among them serving to reduce their number to the figure required by the following theorem *

---

* For a proof of this result see, for instance, E L Ince, "Ordinary Differential Equations," pp 144 150, Dover Publications, Inc , New York, 1944.

THEOREM 1

If the determinant of the operational coefficients of a system of n linear differential equations with constant coefficients is not identically zero, then the number of arbitrary constants in any complete solution of the system is equal to the degree of the determinant of the operational coefficients, regarded as a polynomial in D. In particular cases in which the determinant of the operational coefficients is identically zero, the system may have no solution or it may have solutions containing any number of arbitrary constants

The necessary relations between the constants appearing initially in the solutions for the unknowns can always be found by substituting these solutions into all but one of the n equations in the original system (though not necessarily each set of $n - 1$ equations) and then equating to zero the net coefficients of the terms that occur in each of these equations.

EXAMPLE 1

Find a complete solution of the system

$$(8)\qquad \begin{aligned} (3D^2 + 3D + 2)x + (D^2 + 2D + 3)y &= e^t \\ (2D^2 - D - 2)x + (D^2 + D + 1)y &= 8 \end{aligned}$$

From the preceding discussion we know that the equation satisfied by x is

$$\begin{vmatrix} 3D^2 + 3D + 2 & D^2 + 2D + 3 \\ 2D^2 - D - 2 & D^2 + D + 1 \end{vmatrix} x = \begin{vmatrix} e^t & D^2 + 2D + 3 \\ 8 & D^2 + D + 1 \end{vmatrix}$$

or, expanding the determinants and operating, as required, on the known functions e^t and 8,*

$$(D^4 + 3D^3 + 6D^2 + 12D + 8)x = 3e^t - 24$$

The roots of the characteristic equation of this differential equation are $-1, -2, \pm 2i$ Hence the complementary function is

$$c_1e^{-t} + c_2e^{-2t} + c_3 \cos 2t + c_4 \sin 2t$$

It is easy to see that

$$X = \frac{e^t}{10} - 3$$

is a particular integral, and therefore

$$(9)\qquad x = c_1e^{-t} + c_2e^{-2t} + c_3 \cos 2t + c_4 \sin 2t + \frac{e^t}{10} - 3$$

---

[*] In carrying out these expansions it must be borne in mind that the operational elements in the determinant on the right operate on the algebraic elements e^t and 8, whereas the elements in the determinant on the left *all* operate on x and not on each other This is the reason why in expanding the determinant on the right we have reductions such as

$$D^2 8 = 0 \qquad \text{and} \qquad 2D8 = 0$$

whereas in expanding the determinant on the left we have only formal multiplications such as

$$2D^2 3 = 6D^2 \qquad \text{and} \qquad D^2(-2) = -2D^2$$

The solution for y can now be found by substituting the last expression into either of the original equations and solving the resulting differential equation for y. However, it is usually a little easier to use Cramer's rule again. Doing this, we find that y must satisfy the equation

$$\begin{vmatrix} 3D^2+3D+2 & D^2+2D+3 \\ 2D^2-D-2 & D^2+D+1 \end{vmatrix} y = \begin{vmatrix} 3D^2+3D+2 & e^t \\ 2D^2-D-2 & 8 \end{vmatrix}$$

or

$$(D^4+3D^3+6D^2+12D+8)y = e^t + 16$$

The solution of this presents no difficulty, and we find at once that

$$(10) \qquad y = k_1e^{-t} + k_2e^{-2t} + k_3\cos 2t + k_4\sin 2t + \frac{e^t}{30} + 2$$

However, Eqs. (9) and (10) do not yet constitute the solution of the given system, since collectively they contain eight arbitrary constants, whereas, according to Theorem 1, the complete solution of (8) can contain only four constants. To accomplish the necessary reduction in the number of constants we must now substitute from (9) and (10) into either one or the other (i.e. into all but one) of the original equations, say the second

$$(2D^2-D-2)\left(c_1e^{-t}+c_2e^{-2t}+c_3\cos 2t+c_4\sin 2t+\frac{e^t}{10}-3\right)$$
$$+(D^2+D+1)\left(k_1e^{-t}+k_2e^{-2t}+k_3\cos 2t+k_4\sin 2t+\frac{e^t}{30}+2\right)=8$$

or, performing the indicated differentiations and collecting terms,

$$(c_1+k_1)e^{-t}+(8c_2+3k_2)e^{-2t}+(-10c_3-2c_4-3k_3+2k_4)\cos 2t$$
$$+(2c_3-10c_4-2k_3-3k_4)\sin 2t = 0$$

As it stands, with all eight constants completely arbitrary, this equation is not identically satisfied.* It will be an identity if and only if

$$\begin{aligned} c_1+k_1 &= 0 \\ 8c_2+3k_2 &= 0 \\ -10c_3-2c_4-3k_3+2k_4 &= 0 \\ 2c_3-10c_4-2k_3-3k_4 &= 0 \end{aligned}$$

From these we find (among many equivalent possibilities)

$$k_1 = -c_1 \qquad k_2 = -\tfrac{8}{3}c_2 \qquad k_3 = -2(c_3+c_4) \qquad k_4 = 2(c_3-c_4)$$

Hence, the complete solution to our problem is the pair of functions

$$x = c_1e^{-t}+c_2e^{-2t}+c_3\cos 2t+c_4\sin 2t+\frac{e^t}{10}-3$$

$$y = -c_1e^{-t}-\frac{8}{3}c_2e^{-2t}-2(c_3+c_4)\cos 2t+2(c_3-c_4)\sin 2t+\frac{e^t}{30}+2$$

Though tedious, it is perfectly straightforward to verify that these expressions satisfy the first of the original pair of equations without additional restrictions on the constants.

---

* The reason we encountered no such difficulty in our first illustrative example was that we were able to find x from an equation giving it explicitly in terms of y and its first derivative, and did not have to solve a second differential equation, thereby introducing additional constants.

EXAMPLE 2

Solve the system of equations

$$
\begin{aligned}
Dx + (D-1)y + (D+2)z &= 2e^t \\
(D-1)x + Dy + (D-2)z &= ae^t \\
(D+1)x + (D-2)y + (D+6)z &= e^t
\end{aligned}
$$

From the preceding theory we expect that the differential equation satisfied by z is

$$
\begin{vmatrix} D & D-1 & D+2 \\ D-1 & D & D-2 \\ D+1 & D-2 & D+6 \end{vmatrix} z = \begin{vmatrix} D & D-1 & 2e^t \\ D-1 & D & ae^t \\ D+1 & D-2 & e^t \end{vmatrix}
$$

However, expanding the determinants and operating, as required, on the known functions $2e^t$, ae^t, and e^t, we obtain

$$
0 \cdot z = (a-3)e^t \qquad (10)
$$

Clearly, unless $a = 3$, this equation, and hence the system itself, has no solution On the other hand, if $a = 3$, this equation is satisfied by any function z In fact, if $a = 3$, it is easy to verify that the third equation in the given system is equal to twice the first equation minus the second Hence, when $a = 3$, the last equation is dependent upon the first two and is automatically satisfied by any functions $x(t)$, $y(t)$, $z(t)$ which satisfy them Thus, considering only the first two equations, we can write

$$
\begin{aligned}
Dx + (D-1)y &= 2e^t - (D+2)z \\
(D-1)x + Dy &= 3e^t - (D-2)z
\end{aligned}
$$

and, for every differentiable function z, this system can be solved for x and y Specifically,

$$
\begin{vmatrix} D & D-1 \\ D-1 & D \end{vmatrix} x = \begin{vmatrix} 2e^t - (D+2)z & D-1 \\ 3e^t - (D-2)z & D \end{vmatrix}
$$

or

$$
(2D-1)x = 2e^t - (5D-2)z \tag{11}
$$

and

$$
\begin{vmatrix} D & D-1 \\ D-1 & D \end{vmatrix} y = \begin{vmatrix} D & 2e^t - (D+2)z \\ D-1 & 3e^t - (D-2)z \end{vmatrix}
$$

or

$$
(2D-1)y = 3e^t + (3D-2)z \tag{12}
$$

From Eqs (11) and (12), x and y may be found in terms of z Moreover, since z is subject only to the restriction that it be differentiable, it may contain any number of arbitrary constants, and hence, when $a = 3$, but not otherwise, the solution of the original system may also contain any number of arbitrary constants, as asserted by Theorem 1

EXERCISES

With the understanding that $D \equiv \dfrac{d}{dt}$, find a complete solution of each of the following systems of equations

1 $(D^2+5)x + (D+4)y = e^{-t}$
$(D+2)x + (D+1)y = 3$

2 $(D+5)x + (D+3)y = e^{-t}$
$(D+2)x + (D+1)y = 3$

3 $(D + 5)x + (D + 3)y = e^{-t}$
$(2D + 1)x + (D + 1)y = 3$

4 $(2D + 5)x - (2D + 3)y = t$
$(D - 2)x + (D + 2)y = 0$

5 $(D + 2)x + (D - 1)y = 0$
$(2D + 3)x + (3D + 1)y = 5 \sin 2t$

6 $(D + 1)x + (D + 2)y = e^{t}$
$(2D + 7)x + (D + 10)y = e^{2t}$

7 $(2D + 3)x + (D + 4)y = 0$
$(D + 1)x + (D + 2)y = 0$

8 $(2D^2 + 1)x + (D + 2)y = 5$
$(D^2 - 16)x + (D - 4)y = 4$

9 $(9D^2 + 8)x + (3D^2 + 4)y = 0$
$(2D^2 + 1)x + (D^2 + 2)y = 12 \cos 3t$

10 $(2D^2 - D - 1)x + (D - 1)y = 0$
$(D^2 - 1)x + (D - 1)y = 0$

11
$$\begin{aligned} (D + 1)x + y + 2z &= 1 \\ x + (D + 2)y + z &= e^{-t} + 2 \\ 5x + y + (D - 2)z &= 5e^{-t} + 1 \end{aligned}$$

12
$$\begin{aligned} (D - 1)x - y &= t \\ -2x + (D - 1)y - z &= 0 \\ -2y + (D - 1)z &= e^{2t} \end{aligned}$$

13
$$\begin{aligned} (D + 1)x + (D + 5)y + (2D + 5)z &= 15e^{t} \\ (2D + 1)x + (D + 2)y + (3D + 1)z &= 10e^{t} \\ (D + 3)x + (3D + 4)y + (4D + 6)z &= 21e^{t} \end{aligned}$$

14
$$\begin{aligned} (D + 1)x + (D + 3)y + (2D + 3)z &= e^{t} \\ (2D + 1)x + (D + 2)y + (3D + 1)z &= 0 \\ (D + 3)x + (3D + 11)y + (4D + 13)z &= 0 \end{aligned}$$

15
$$\begin{aligned} (D + 1)x + (D + 1)y + (2D + 3)z &= 0 \\ (2D + 1)x + (D + 2)y + (3D + 5)z &= 0 \end{aligned}$$
$(D + 3)x + (3D + 1)y + ($[illegible]$ + 5)z = 0$

16 If (x_1,y_1) and (x_2,y_2) are two solutions of the system

$$\begin{aligned} P_{11}(D)x + P_{12}(D)y &= 0 \\ P_{21}(D)x + P_{22}(D)y &= 0 \end{aligned}$$

prove that $(c_1x_1 + c_2x_2,\ c_1y_1 + c_2y_2)$ is also a solution of this system

17 Find a system of differential equations having

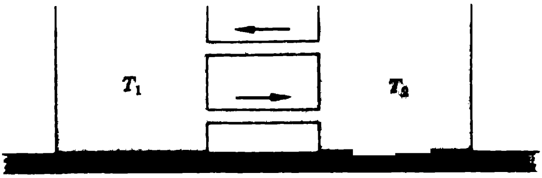
$$\begin{aligned} x &= Ae^{-t} + Be^{t} + Ce^{2t} \\ y &= Ae^{-t} - Be^{t} + 2Ce^{2t} \end{aligned}$$

as a complete solution

18 In Example 1, determine multiples of the two equations which, when added, will yield an equation expressing y directly in terms of x and its various derivatives Do you think that this can be done in general?

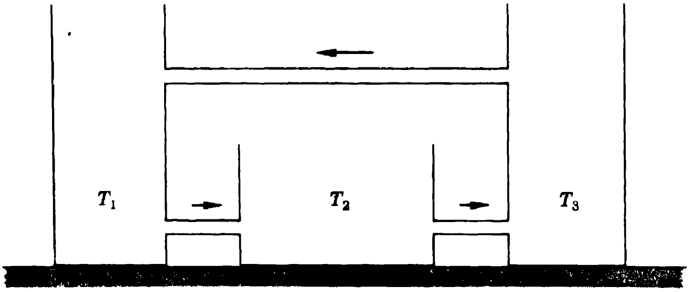
19 Two tanks are connected as shown in Fig 3 1 The first tank contains 100 gal of pure water, the second contains 100 gal of brine containing 2 lb of salt per gal Liquid circulates through the tanks at a constant rate of 5 gal per min If the brine in each tank is kept uniform by stirring, find the amounts of salt in the respective tanks as functions of time

FIGURE 3 1



20 Three tanks are connected as shown in Fig 3 2 The first tank contains 100 gal of pure water, the second contains 100 gal of brine containing 1 lb of salt per gal, the third contains 100 gal of brine containing 2 lb of salt per gal. Liquid circulates through the tank at a constant rate of 5 gal per min If the brine in each tank is kept uniform by stirring, find the amounts of salt in the respective tanks as functions of time.

FIGURE 3.2



3.3

Complementary functions and particular integrals for systems of equations

To illustrate the extension of the ideas of *characteristic equation, complementary function,* and *particular integral* to systems of differential equations, let us consider the following set of equations:

$$
(1)\qquad
\begin{aligned}
(D+1)x + (D+2)y + (D+3)z &= -e^{-t} + 8t + 2\\
(D+2)x + (D+3)y + (2D+3)z &= e^{-t} + 11t - 1\\
(4D+6)x + (5D+4)y + (20D-12)z &= 7e^{-t} + 2t
\end{aligned}
$$

As in the case of a single equation, we shall first make the system homogeneous by neglecting the terms on the right, getting

$$
(2)\qquad
\begin{aligned}
(D+1)x + (D+2)y + (D+3)z &= 0\\
(D+2)x + (D+3)y + (2D+3)z &= 0\\
(4D+6)x + (5D+4)y + (20D-12)z &= 0
\end{aligned}
$$

Guided by our experience in solving single equations, let us now attempt to find solutions of this system of the form

$$
(3)\qquad x = ae^{mt} \qquad y = be^{mt} \qquad z = ce^{mt}
$$

Substituting these into the equations in (2) and dividing out the common factor e^{mt} leads to the set of algebraic equations

$$
(4)\qquad
\begin{aligned}
(m+1)a + (m+2)b + (m+3)c &= 0\\
(m+2)a + (m+3)b + (2m+3)c &= 0\\
(4m+6)a + (5m+4)b + (20m-12)c &= 0
\end{aligned}
$$

If nontrivial solutions for x, y, and z, i.e , solutions that do not vanish identically, are to be obtained, it is necessary that a, b, and c shall not all be zero. However, the values $a = b = c = 0$ obviously satisfy the system (4) and in general will be the only solution of this set of equations In fact, from college algebra (or from Corollary 1, Theorem 7, Sec 10 5) we know that no other solutions are possible unless the determinant of the coefficients in (4) is equal to zero Thus we must have

$$\begin{vmatrix} m+1 & m+2 & m+3 \\ m+2 & m+3 & 2m+3 \\ 4m+6 & 5m+4 & 20m-12 \end{vmatrix} = -(m-1)(m-2)(m-3) = 0$$

This equation, which defines all the values of m for which nontrivial solutions of (4), and hence of (2), can exist, is the **characteristic equation** of the system It is, of course, nothing but the determinant of the operational coefficients of the system equated to zero, with D replaced by m

From the roots of this equation, $m_1 = 1$, $m_2 = 2$, $m_3 = 3$, we can construct three particular solutions

$$(5) \qquad \begin{cases} x_1 = a_1e^t \\ y_1 = b_1e^t \\ z_1 = c_1e^t \end{cases} \qquad \begin{cases} x_2 = a_2e^{2t} \\ y_2 = b_2e^{2t} \\ z_2 = c_2e^{2t} \end{cases} \qquad \begin{cases} x_3 = a_3e^{3t} \\ y_3 = b_3e^{3t} \\ z_3 = c_3e^{3t} \end{cases}$$

provided that we establish the proper relations among the constants in each of the three sets

To do this, we note that the constants a_i, b_i, c_i must satisfy the equations of the system (4) for the corresponding value m_i Thus for $m_1 = 1$ we must have

$$2a_1 + 3b_1 + 4c_1 = 0$$
$$3a_1 + 4b_1 + 5c_1 = 0$$
$$10a_1 + 9b_1 + 8c_1 = 0$$

We know, of course, that the determinant of the coefficients of this system is equal to zero Hence these equations are consistent, i e , have a solution other than $a_1 = b_1 = c_1 = 0$, and it is easy to verify that, for all values of k_1, they are satisfied by

$$a_1 = -k_1 \qquad b_1 = 2k_1 \qquad c_1 = -k_1$$

Therefore, for each value of k_1,

$$(6) \qquad \begin{aligned} x_1 &= -k_1e^t \\ y_1 &= 2k_1e^t \\ z_1 &= -k_1e^t \end{aligned}$$

is a particular solution of (2) corresponding to the characteristic root $m_1 = 1$.

Similarly, for $m_2 = 2$, we have from (4)

$$\begin{aligned} 3a_2 + 4b_2 + 5c_2 &= 0 \\ 4a_2 + 5b_2 + 7c_2 &= 0 \\ 14a_2 + 14b_2 + 28c_2 &= 0 \end{aligned}$$

and it is easy to verify that, for all values of k_2, these are satisfied by

$$a_2 = 3k_2 \qquad b_2 = -k_2 \qquad c_2 = -k_2$$

Therefore, a second family of particular solutions of (2) is

$$\begin{aligned} x_2 &= 3k_2e^{2t} \\ y_2 &= -k_2e^{2t} \\ z_2 &= -k_2e^{2t} \end{aligned} \tag{7}$$

Finally, for $m_3 = 3$, we have from (4)

$$\begin{aligned} 4a_3 + 5b_3 + 6c_3 &= 0 \\ 5a_3 + 6b_3 + 9c_3 &= 0 \\ 18a_3 + 19b_3 + 48c_3 &= 0 \end{aligned}$$

and $\quad a_3 = 9k_3 \qquad b_3 = -6k_3 \qquad c_3 = -k_3$

A third family of particular solutions of (2) is, therefore,

$$\begin{aligned} x_3 &= 9k_3e^{3t} \\ y_3 &= -6k_3e^{3t} \\ z_3 &= -k_3e^{3t} \end{aligned} \tag{8}$$

Since the equations of the homogeneous system (2) are all linear, sums of solutions will also be solutions Hence we can combine the three families of particular solutions (6), (7), and (8) into a complete solution of (2).

$$\begin{aligned} x &= x_1 + x_2 + x_3 = -k_1e^t + 3k_2e^{2t} + 9k_3e^{3t} \\ y &= y_1 + y_2 + y_3 = 2k_1e^t - k_2e^{2t} - 6k_3e^{3t} \\ z &= z_1 + z_2 + z_3 = -k_1e^t - k_2e^{2t} - k_3e^{3t} \end{aligned} \tag{9}$$

This is the **complementary function** of the original nonhomogeneous system (1) We note that it contains precisely three arbitrary constants, as required by Theorem 1, Sec 3 2 The relations among the nine constants originally present in the three particular solutions (5) could also have been found by substituting those solutions into any two of the equations of the homogeneous system (2) and equating coefficients, as we did in Example 1, Sec 3.2

To complete the problem we now need to find a particular solution, or "integral," of the nonhomogeneous system (1) To do this, we assume for x, y, and z individual trial solutions exactly as described in Table 2.2, Sec. 2.3. Thus, in the present case we choose

$$X = \alpha_1 e^{-t} + \alpha_2 t + \alpha_3 \qquad Y = \beta_1 e^{-t} + \beta_2 t + \beta_3 \qquad Z = \gamma_1 e^{-t} + \gamma_2 t + \gamma_3$$

Substituting these into (1) and collecting terms, we find

$$(\beta_1 + 2\gamma_1)e^{-t} + (\alpha_2 + 2\beta_2 + 3\gamma_2)t + (\alpha_2 + \beta_2 + \gamma_2 + \alpha_3 + 2\beta_3 + 3\gamma_3) = -e^{-t} + 8t + 2$$

$$(\alpha_1 + 2\beta_1 + \gamma_1)e^{-t} + (2\alpha_2 + 3\beta_2 + 3\gamma_2)t + (\alpha_2 + \beta_2 + 2\gamma_2 + 2\alpha_3 + 3\beta_3 + 3\gamma_3) = e^{-t} + 11t - 1$$

$$(2\alpha_1 - \beta_1 - 32\gamma_1)e^{-t} + (6\alpha_2 + 4\beta_2 - 12\gamma_2)t + (4\alpha_2 + 5\beta_2 + 20\gamma_2 + 6\alpha_3 + 4\beta_3 - 12\gamma_3) = 7e^{-t} + 2t$$

Clearly, these three equations will hold identically if and only if the following sets of conditions are satisfied

$$\begin{aligned} \beta_1 + 2\gamma_1 &= -1 \\ \alpha_1 + 2\beta_1 + \gamma_1 &= 1 \\ 2\alpha_1 - \beta_1 - 32\gamma_1 &= 7 \end{aligned} \tag{10}$$

$$\begin{aligned} \alpha_2 + 2\beta_2 + 3\gamma_2 &= 8 \\ 2\alpha_2 + 3\beta_2 + 3\gamma_2 &= 11 \\ 6\alpha_2 + 4\beta_2 - 12\gamma_2 &= 2 \end{aligned} \tag{11}$$

$$\begin{aligned} \alpha_2 + \beta_2 + \gamma_2 + \alpha_3 + 2\beta_3 + 3\gamma_3 &= 2 \\ \alpha_2 + \beta_2 + 2\gamma_2 + 2\alpha_3 + 3\beta_3 + 3\gamma_3 &= -1 \\ 4\alpha_2 + 5\beta_2 + 20\gamma_2 + 6\alpha_3 + 4\beta_3 - 12\gamma_3 &= 0 \end{aligned} \tag{12}$$

From the set (10) we find without difficulty that

$$\alpha_1 = 3 \qquad \beta_1 = -1 \qquad \gamma_1 = 0$$

From (11) we find that

$$\alpha_2 = 1 \qquad \beta_2 = 2 \qquad \gamma_2 = 1$$

Finally, from (12), after the values for α_2, β_2, and γ_2 are inserted, we find that

$$\alpha_3 = -3 \qquad \beta_3 = -1 \qquad \gamma_3 = 1$$

With these values for the constants, the particular integral of the nonhomogeneous system (1) becomes

$$X = 3e^{-t} + t - 3 \qquad Y = -e^{-t} + 2t - 1 \qquad Z = t + 1$$

Hence, adding these to the respective components of the complementary function (9), we have the complete solution of the original system

$$\begin{aligned} x &= -k_1e^t + 3k_2e^{2t} + 9k_3e^{3t} + 3e^{-t} + t - 3 \\ y &= 2k_1e^t - k_2e^{2t} - 6k_3e^{3t} - e^{-t} + 2t - 1 \\ z &= -k_1e^t - k_2e^{2t} - k_3e^{3t} + t + 1 \end{aligned}$$

EXERCISES

Find a complete solution of each of the following systems

1 $(D + 2)x + (D + 4)y = 1$
$(D + 1)x + (D + 5)y = 2$

2 $(2D + 1)x + (D + 2)y = 0$
$(D + 3)x + (D + 6)y = -3e^t$

3 $(D + 1)x + (4D - 2)y = t - 1$
$(D + 2)x + (5D - 2)y = 2t - 1$

4 $(D + 5)x + (D + 7)y = 8e^{2t}$
$(2D + 1)x + (3D + 1)y = 0$

5 $(2D + 1)x + (D + 2)y = e^t$
$(D + 2)x + (D + 4)y = e^{-t}$

6 $(2D + 1)x + (D - 1)y = \cos t$
$(D + 2)x + (D + 3)y = 0$

7 $(2D^2 + 5)x + (D^2 + 3)y = 1$
$(D^2 + 7)x + (D^2 + 5)y = t$ (Hint Assume first $x = a \cos \lambda t$, $y = b \cos \lambda t$, and then $x = c \sin \lambda t$, $y = d \sin \lambda t$, where λ is a parameter to be determined)

8 $(2D^2 + 7)x + (D^2 + 5)y = e^{-t}$
$(3D^2 + 13)x + (3D^2 + 11)y = 2e^{-t} + 12$

9 $(2D^2 + 15)x + (D^2 + 12)y = \cos t$
$(3D^2 + 26)x + (3D^2 + 28)y = 0$

10
$$\begin{aligned}(2D + 11)x + (D + 3)y + (D - 2)z &= 14e^t\\(D - 2)x + (D - 1)y + Dz &= -2e^t\\(D + 1)x + (D - 3)y + (2D - 4)z &= 4e^t\end{aligned}$$

CHAPTER FOUR

Finite Differences

4.1

The differences of a function

In the last three chapters we have developed methods for the solution of several large and important classes of differential equations There are, of course, other families of equations for which exact solutions can be found, but in general, differential equations more complicated than the simple ones we have been considering must be solved by approximate, numerical methods Among the most important of these are what are known as *finite-difference methods* Since finite differences also occur in other branches of numerical analysis, such as interpolation, numerical differentiation and integration, curve fitting, and the smoothing of data, it is desirable that an applied mathematician have some familiarity with them, and the present chapter is devoted to this end

Suppose that we have a function $y = f(x)$ given in tabular form for a sequence of values of x

x	$f(x)$
x_0	$f(x_0)$
x_1	$f(x_1)$
x_2	$f(x_2)$
x_3	$f(x_3)$

If $f(x_i)$ and $f(x_j)$ are any two values of $f(x)$, then the **first divided differences** of $f(x)$ are defined by the formula*

(1) $$f(x_i,x_j) = \frac{f(x_i) - f(x_j)}{x_i - x_j}$$

Similarly, if $f(x_i,x_j)$ and $f(x_j,x_k)$ are two first differences of $f(x)$

---

* In most applications the subscripts of the arguments x_i and x_j will be consecutive integers, but this is not a necessary restriction on the definition

having one argument, x_j, in common, then the **second divided differences** of $f(x)$ are defined by the formula

$$f(x_i,x_j,x_k) = \frac{f(x_i,x_j) - f(x_j,x_k)}{x_i - x_k} \tag{2}$$

Proceeding inductively, a divided difference of any order is defined as the difference between two divided differences of the next lower order, overlapping in all but one of their arguments, divided by the difference between the extreme, or nonoverlapping, arguments appearing in these differences * From these definitions it is clear that divided differences have the following properties

PROPERTY 1

Any divided difference of the sum (or difference) of two functions is equal to the sum (or difference) of the divided differences of the individual functions

PROPERTY 2

Any divided difference of a constant times a function is equal to the constant times the divided difference of the function

In many applications it is convenient to have the divided differences of a function prominently displayed This is usually done by constructing a **difference table** in which each difference is entered, in the appropriate column, midway between the elements in the preceding column from which it is constructed

x	$f(x)$			
x_0	$f(x_0)$			
		$f(x_0,x_1)$		
x_1	$f(x_1)$		$f(x_0,x_1,x_2)$	
		$f(x_1,x_2)$		$f(x_0,x_1,x_2,x_3)$
x_2	$f(x_2)$		$f(x_1,x_2,x_3)$	
		$f(x_2,x_3)$		
x_3	$f(x_3)$			

or in a specific numerical example,

x	x^3				
0	0				
		1			
1	1		4		
		13		1	
3	27		8		0
		37		1	
4	64		14		0
		93		1	
7	343		20		
		193			
9	729				

---

* Though obvious only for divided differences of the first order, it is true (see Exercises 13 and 14) that divided differences of all orders are symmetric functions of their arguments Thus, $f(x_i,x_j,x_k) = f(x_i,x_k,x_j) = f(x_j,x_i,x_k) =$

Usually the values of x in a table of data will be equally spaced, and the differences of the function will be based on sets of consecutive functional values. When this is the case, the denominators in the divided differences of any given order are all the same, and it is customary to omit them. This leads to a modified set of quantities known simply as the **differences** of the function. If the constant difference between successive values of x is h, so that the general value of x in the table is

$$x_k = x_0 + kh \qquad k = \ldots, -2, -1, 0, 1, 2, \ldots$$

and the corresponding functional value is

$$y_k = f(x_k) = f(x_0 + kh) = f_k$$

then the **first differences** of f are defined by the formula

(3) $$\Delta f_k = f_{k+1} - f_k$$

Differences of higher order are defined in the same way, the **second differences** being

(4) $$\Delta^2 f_k = \Delta(\Delta f_k) = \Delta f_{k+1} - \Delta f_k$$

and, in general, for positive integral values of n,

(5) $$\Delta^n f_k = \Delta(\Delta^{n-1} f_k) = \Delta^{n-1} f_{k+1} - \Delta^{n-1} f_k$$

These differences are also displayed in difference tables just like divided differences.

Evidently the **difference operator** Δ has the characteristic properties of a linear operator, for

$$\begin{aligned}\Delta(f_k \pm g_k) &= (f_{k+1} \pm g_{k+1}) - (f_k \pm g_k)\\ &= (f_{k+1} - f_k) \pm (g_{k+1} - g_k) = \Delta f_k \pm \Delta g_k\end{aligned}$$

and, if c is a constant,

$$\Delta(cf_k) = cf_{k+1} - cf_k = c(f_{k+1} - f_k) = c\,\Delta f_k$$

Moreover, Δ obeys the usual law of exponents

$$\Delta^m(\Delta^n f_k) = \Delta^{m+n} f_k$$

provided both m and n are positive integers.

When the values of the independent variable are equally spaced, the divided differences of a function can easily be expressed in terms of ordinary differences and vice versa. Specifically,

$$f(x_0, x_1) = \frac{f(x_0) - f(x_1)}{x_0 - x_1} = \frac{f_0 - f_1}{-h} = \frac{\Delta f_0}{h}$$

$$f(x_0, x_1, x_2) = \frac{f(x_0, x_1) - f(x_1, x_2)}{x_0 - x_2} = -\frac{1}{2h}\left(\frac{\Delta f_0}{h} - \frac{\Delta f_1}{h}\right) = \frac{\Delta^2 f_0}{2!h^2}$$

and, in general,

(6) $$f(x_0, x_1, \ldots, x_n) = \frac{\Delta^n f_0}{n!h^n}$$

More generally, if the points used in constructing an nth divided difference are the $n + 1$ equally spaced points between $x_0 - kh$

and $x_0 + (n - k)h$, inclusive, it is easy to show that

(7) $$f(x_{-k}, x_{-k+1}, \ldots, x_{n-k}) = \frac{\Delta^n f_{-k}}{n!\,h^n}$$

The Δ symbolism for the differences of a function is known as the **advancing difference notation.** In some applications, however, another notation known as the **central difference notation** is more convenient In this, the symbol δ is used instead of Δ, and the subscript appearing in the symbol for any difference is the average of the subscripts already assigned by this convention to the elements which are subtracted in forming that difference Thus,

$$\Delta f_k = f_{k+1} - f_k = \delta f_{k+\frac{1}{2}} \qquad \Delta f_{k+1} = f_{k+2} - f_{k+1} = \delta f_{k+\frac{3}{2}}$$

$$\Delta^2 f_k = \Delta f_{k+1} - \Delta f_k = \delta f_{k+\frac{3}{2}} - \delta f_{k+\frac{1}{2}} = \delta^2 f_{k+1}$$

$$\cdot \qquad \cdot$$

The following difference tables show the relation between the advancing and the central difference notations

x	f				
x_0	f_0				
		Δf_0			
x_1	f_1		$\Delta^2 f_0$		
		Δf_1		$\Delta^3 f_0$	
x_2	f_2		$\Delta^2 f_1$		$\Delta^4 f_0$
		Δf_2		$\Delta^3 f_1$	
x_3	f_3		$\Delta^2 f_2$		
		Δf_3			
x_4	f_4				

x	f				
x_0	f_0				
		$\delta f_{\frac{1}{2}}$			
x_1	f_1		$\delta^2 f_1$		
		$\delta f_{\frac{3}{2}}$		$\delta^3 f_{\frac{3}{2}}$	
x_2	f_2		$\delta^2 f_2$		$\delta^4 f_2$
		$\delta f_{\frac{5}{2}}$		$\delta^3 f_{\frac{5}{2}}$	
x_3	f_3		$\delta^2 f_3$		
		$\delta f_{\frac{7}{2}}$			
x_4	f_4				

In the first, elements with the same subscript lie on lines sloping downward, or *advancing* into the table In the second, elements with the same subscript lie on lines extending horizontally, or *centrally*, into the table

Closely associated with Δ and δ is the operator E, which is defined as the operator which increases the argument of a function by one tabular interval Thus,

$$Ef(x_k) = f(x_k + h) = f(x_{k+1}) = f_{k+1}$$

Applying E a second time again increases the argument of f by h; that is,

$$E^2 f(x_k) = E[Ef(x_k)] = Ef(x_k + h) = f(x_k + 2h) = f(x_{k+2}) = f_{k+2}$$

and, in general, we define

(8) $$E^r f(x_k) = f(x_k + rh) = f(x_{k+r}) = f_{k+r}$$

for any real number r Clearly, E obeys the laws

$$E(f_k \pm g_k) = Ef_k \pm Eg_k$$

$$E(cf_k) = cEf_k \qquad c \text{ a constant}$$

$$E^r(E^s f_k) = E^{r+s} f_k$$

Two operators with the property that when they are applied to the same function they yield the same result are said to be **operationally equivalent.** Now, from the definition of Δf_k, we have

$$\Delta f_k = f_{k+1} - f_k = Ef_k - f_k$$

or, symbolically, $\Delta f_k = (E - 1)f_k$

Hence, we have the operational equivalences

$$\Delta = E - 1 \tag{9}$$

$$E = 1 + \Delta \tag{10}$$

$$E - \Delta = 1 \tag{11}$$

Moreover, by definition,

$$\Delta f_k = \delta f_{k+\frac{1}{2}} = \delta E^{\frac{1}{2}} f_k$$

Hence, we have the further equivalences

$$\Delta = \delta E^{\frac{1}{2}} \tag{12}$$

$$\delta = \Delta E^{-\frac{1}{2}} \tag{13}$$

Also, substituting from (12) into (9) and solving for δ, we have

$$\delta = E^{\frac{1}{2}} - E^{-\frac{1}{2}} \tag{14}$$

By means of (9) we can express the various differences of a function in terms of successive entries in the table of the function. For we can write

$$\Delta^n f_k = (E - 1)^n f_k$$

and then, using the binomial expansion,

$$\begin{aligned}\Delta^n f_k &= \left[E^n - \binom{n}{1} E^{n-1} + \binom{n}{2} E^{n-2} + \right.\\ &\qquad \left. + (-1)^{n-1}\binom{n}{n-1} E + (-1)^n \binom{n}{n}\right] f_k\dagger \\ &= E^n f_k - nE^{n-1} f_k + \frac{n(n-1)}{2!} E^{n-2} f_k + \\ &\qquad + (-1)^{n-1} nEf_k + (-1)^n f_k \\ &= f_{k+n} - nf_{k+n-1} + \frac{n(n-1)}{2!} f_{k+n-2} + \\ &\qquad + (-1)^{n-1} nf_{k+1} + (-1)^n f_k \end{aligned} \tag{15}$$

Specifically, taking $k = 0$ and $n = 1, 2, 3, 4,$, we have

$$\begin{aligned} \Delta f_0 &= f_1 - f_0 \\ \Delta^2 f_0 &= f_2 - 2f_1 + f_0 \\ \Delta^3 f_0 &= f_3 - 3f_2 + 3f_1 - f_0 \\ \Delta^4 f_0 &= f_4 - 4f_3 + 6f_2 - 4f_1 + f_0 \end{aligned} \tag{15a}$$

. . .

---

† The quantities $\binom{n}{j}$ are the so-called **binomial coefficients**, defined by the formula $\binom{n}{j} = \dfrac{n!}{j!(n-j)!}$

The fact that the first divided difference of a function is precisely the difference quotient whose limit defines the derivative of the function suggests that in some respects the properties of the differences of a function and the properties of the derivatives of a function may be analogous This is actually the case, and among other interesting results we have the following·

THEOREM 1

The nth divided differences of a polynomial of degree n are constant

PROOF To prove this theorem, it is clearly sufficient to establish the asserted property for the special polynomial x^n To do this, we observe that for x^n the first divided difference is simply

$$f(x_i,x_j) = \frac{x_i^n - x_j^n}{x_i - x_j} = x_i^{n-1} + x_i^{n-2}x_j + \cdots + x_ix_j^{n-2} + x_j^{n-1}$$

which is a homogeneous and symmetric function of x_i and x_j of degree $n - 1$ For the second divided difference we have, of course,

$$f(x_i,x_j,x_k) = \frac{f(x_i,x_j) - f(x_j,x_k)}{x_i - x_k}$$

Moreover, since divided differences of all orders are symmetric functions of their arguments (see the footnote on page 80), it follows that the numerator of the last fraction vanishes when $x_i = x_k$ Hence, it must contain $x_i - x_k$ as a factor, and, therefore, as we verified explicitly for the first divided difference, the indicated division is exact Thus the second divided difference of x^n is a homogeneous and symmetric expression of degree $n - 2$ in x_i, x_j, and x_k Continuing in this way, it is evident that, after differencing n times, the degree of the resultant expression will be zero. that is, x^n will have been reduced to a constant independent of $x_i, x_j, x_k, \ldots$, as asserted

Since ordinary differences are proportional to the corresponding divided differences, it is clear that Theorem 1 also holds for these differences whether they are expressed in terms of the advancing or the central difference notation. Thus we have the following corollary of Theorem 1·

COROLLARY 1

The nth (ordinary) differences of a polynomial of degree n are constant

The analogy between differences and derivatives becomes even more striking if we consider the operator Δ and introduce the so-called **factorial polynomials**:

(16) $$(x)^{(n)} = x(x - 1) \cdots (x - n + 1)$$

(17) $$(x)^{-(n)} = \frac{1}{(x + 1)(x + 2) \cdots (x + n)}$$

In general, these play the same role in the calculus of finite differences that the power functions x^n and x^{-n} play in ordinary calculus In particular, we have the important formulas

(18) $$\Delta(x)^{(n)} = n(x)^{(n-1)}$$

(19) $$\Delta(x)^{-(n)} = -n(x)^{-(n+1)}$$

whose resemblance to the formulas for differentiating x^n and x^{-n} is unmistakable. The proofs of these involve only a little elementary algebra, and we shall leave them as exercises.

In view of Formulas (18) and (19), it is a matter of some interest to be able to express an arbitrary polynomial $p(x)$ of degree n in terms of factorial polynomials. One way this can be done is to write, by analogy with Maclaurin's expansion,

(20) $$p(x) = a_0 + a_1(x)^{(1)} + a_2(x)^{(2)} + \cdots + a_n(x)^{(n)}$$

where, clearly, since $p(x)$ is of degree n, no terms beyond $a_n(x)^{(n)}$ need be included. If we set $x = 0$ in Eq. (20), every term after the first becomes zero, since each contains x as a factor. Hence,

$$a_0 = p(0)$$

Now if we use (18) to take the first difference of $p(x)$, as given by Eq. (20), we get

$$\Delta p(x) = a_1 + 2a_2(x)^{(1)} + 3a_3(x)^{(2)} + \cdots + na_n(x)^{(n-1)}$$

and, if we set $x = 0$ in this expression, we obtain

$$a_1 = \Delta p(0)$$

Differencing again, we find

$$\Delta^2 p(x) = 2!a_2 + 3\cdot 2a_3(x)^{(1)} + \cdots + n(n-1)a_n(x)^{(n-2)}$$

and, evaluating at $x = 0$,

$$a_2 = \frac{\Delta^2 p(0)}{2!}$$

Continuing this process of differencing and evaluating at $x = 0$, we obtain the general formula

(21) $$a_j = \frac{\Delta^j p(0)}{j!} \qquad j = 0, 1, \ldots, n$$

which obviously resembles closely the familiar formula for the coefficients in Maclaurin's expansion. This procedure is especially convenient when we are given a difference table of the polynomial rather than the polynomial itself.

When we are given the polynomial itself, it is usually inefficient to construct a difference table and apply the preceding method. Instead, it is better to proceed in the following way. If we divide $p(x)$ by x we get a remainder r_0 (which is just the constant term in p) and a quotient $q_0(x)$, so that we can write

(22) $$p(x) = r_0 + xq_0(x)$$

Now, if we divide $q_0(x)$ by $x - 1$, we get a remainder r_1 and a quotient $q_1(x)$ such that

$$q_0(x) = r_1 + (x - 1)q_1(x)$$

Hence, substituting into (22),

(23) $$p(x) = r_0 + x[r_1 + (x - 1)q_1(x)] = r_0 + r_1(x)^{(1)} + x(x - 1)q_1(x)$$

If, further, we divide $q_1(x)$ by $x - 2$, we obtain a remainder r_2 and a quotient $q_2(x)$ such that

$$q_1(x) = r_2 + (x - 2)q_2(x)$$

and, substituting into (23),

$$\begin{aligned} p(x) &= r_0 + r_1(x)^{(1)} + x(x - 1)[r_2 + (x - 2)q_2(x)] \\ &= r_0 + r_1(x)^{(1)} + r_2(x)^{(2)} + x(x - 1)(x - 2)q_2(x) \end{aligned}$$

Each application of this procedure leads to a new quotient whose degree is one less than the degree of the preceding quotient. Hence, the process must terminate after $n + 1$ steps with the required expansion

(24) $$p(x) = r_0 + r_1(x)^{(1)} + r_2(x)^{(2)} + \cdots + r_{n-1}(x)^{(n-1)} + r_n(x)^{(n)}$$

Obviously, the required divisions can easily be carried out by the elementary process of synthetic division. Moreover, it is clear from Eqs. (20), (21), and (24) that

$$r_j = a_j = \frac{\Delta^j p(0)}{j!}$$

or

(25) $$\Delta^j p(0) = j!\, r_j$$

Hence, this method provides a convenient way of constructing the difference table of a polynomial in the important case when $h = 1$, since it furnishes us with the leading entry in each column of the table and from these the table can be extended as far as desired by simple addition, using the identity

$$\Delta^{j-1} f_{k+1} = \Delta^{j-1} f_k + \Delta^j f_k$$

EXAMPLE 1

Express $p(x) = x^4 - 5x^3 + 3x + 4$ in terms of factorial polynomials and construct the difference table of the function for $h = 1$.

Using synthetic division we have at once

$$\begin{array}{r|rrrrr}
\underline{1} & 1 & -5 & 0 & 3 & 4 \\
 & & 1 & -4 & -4 & \\ \hline
\underline{2} & 1 & -4 & -4 & \underline{-1} & \\
 & & 2 & -4 & & \\ \hline
\underline{3} & 1 & -2 & \underline{-8} & & \\
 & & 3 & & & \\ \hline
 & 1 & 1 & & &
\end{array}$$

The remainders r_0, r_1, r_2, r_3, r_4 are the underscored numbers 4, -1, -8, 1, 1. Hence

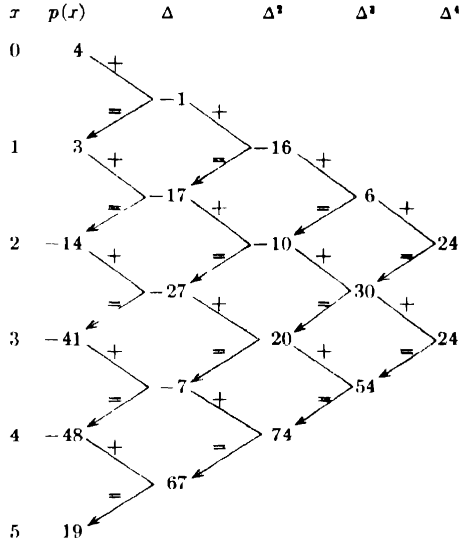
$$p(x) \equiv x^4 - 5x^3 + 3x + 4 \equiv 4 - (x)^{(1)} - 8(x)^{(2)} + (x)^{(3)} + (x)^{(4)}$$

as can be verified by direct expansion.

Now from (25) we have

$$p(0) = 4 \qquad \Delta p(0) = -1 \qquad \Delta^2 p(0) = -16 \qquad \Delta^3 p(0) = 6 \qquad \Delta^4 p(0) = 24$$

Hence we have the leading entries in the difference table for $p(x)$, and by "crisscross" addition, as indicated, the table can be extended and the values of $p(x)$ determined as far as may be desired



Once a function has been expressed as a series of factorial polynomials, it is a simple matter to apply Eq. (18) or (19) to obtain its various differences Conversely, when a function has been expressed as a series of factorial polynomials, it is easy to use these equations "in reverse" and find a new function having the given function as its first difference By analogy with the terminology of calculus, we shall refer to such a function as an **antidifference.**

EXAMPLE 2

What is the general antidifference of the polynomial $p(x) = x^4 - 5x^3 + 3x + 4$?

From the results of Example 1 we know that

$$p(x) = (x)^{(4)} + (x)^{(3)} - 8(x)^{(2)} - (x)^{(1)} + 4$$

Hence, from Eq (18), it is clear that the required antidifference, which is often denoted by the symbol $\Delta^{-1}p(x)$, is

$$\Delta^{-1}p(x) = \frac{(x)^{(5)}}{5} + \frac{(x)^{(4)}}{4} - \frac{8(x)^{(3)}}{3} - \frac{(x)^{(2)}}{2} + 4(x)^{(1)} + c$$

where c is an arbitrary constant which can, and in general must, be added, since the difference of any constant is obviously zero The analogy between antidifferences and indefinite integrals or antiderivatives is clear

The determination of antidifferences is not just a mathematical curiosity, but is intimately related to the important problem of finding the sums of series To see this, consider any two consecutive columns in a difference table

$$\begin{array}{ll}
\Delta^k f_1 & \\
& \Delta^{k+1} f_1 \\
\Delta^k f_2 & \\
& \Delta^{k+1} f_2 \\
& \\
& \\
\Delta^k f_n & \\
& \Delta^{k+1} f_n \\
\Delta^k f_{n+1} &
\end{array}$$

Now, from the definition of a difference we have

$$\sum_{i=1}^{n} \Delta^{k+1} f_i = (\Delta^k f_2 - \Delta^k f_1) + (\Delta^k f_3 - \Delta^k f_2) + \cdots + (\Delta^k f_n - \Delta^k f_{n-1}) + (\Delta^k f_{n+1} - \Delta^k f_n)$$

or, canceling the common terms in the series on the right,

(26) $$\sum_{i=1}^{n} \Delta^{k+1} f_i = \Delta^k f_{n+1} - \Delta^k f_1$$

Since the kth difference of a function is obviously an antidifference of the $(k + 1)$st difference, it is clear that Eq (26) is equivalent to the following theorem

THEOREM 2

If $F(i)$ is any antidifference of $f(i)$, then the sum from $i = 1$ to $i = n$ of the series whose general term is $f(i)$ is $F(n + 1) - F(1)$

The analogy between this theorem and the fundamental theorem of integral calculus is unmistakable

EXAMPLE 3

What is the sum of the squares of the first n odd integers?

To facilitate the finding of the necessary antidifference, we first express the general term of the series, namely, $(2i - 1)^2$, in terms of factorial polynomials

$$(2i - 1)^2 = 4i(i - 1) + 1 = 4(i)^{(2)} + 1$$

Then, by the last theorem,

$$\sum_{i=1}^{n} (2i - 1)^2 = \sum_{i=1}^{n} [4(i)^{(2)} + 1] = \left[\frac{4(i)^{(3)}}{3} + (i)^{(1)}\right]_{i=1}^{i=n+1}$$

$$= \frac{4(n + 1)^{(3)}}{3} + (n + 1)^{(1)} - \frac{4(1)^{(3)}}{3} - (1)^{(1)}$$

$$= \frac{4(n + 1)n(n - 1)}{3} + (n + 1) - 0 - 1$$

$$4n^3 - n$$

EXERCISES

1 Prove Formulas (6) and (7) **2** Prove Formulas (18) and (19)

3 Express the following polynomials in terms of factorial polynomials, and construct a difference table for each function

a $x^3 - x + 1$ **b** $x^4 - 2x^3 - x$ **c** $x^5 - 2x^4 + 4x^3 - x + 6$

4 For each of the following difference tables, find the polynomial of minimum degree which yields the given data.

a

x	y				
0	-1				
		-1			
1	-2		6		
		5		6	
2	3		12		0
		17		6	
3	20		18		
		35			
4	55				

b

x	y					
0	6					
		-5				
1	1		2			
		-3		-6		
2	-2		-4		24	
		-7		18		0
3	-9		14		24	
		7		42		
4	-2		56			
		63				
5	61					

5 If $h = 1$, show that, for all values of the constants a and b, each of the following functions satisfies the indicated relation

a $y = a2^x + b3^x$ $(E^2 - 5E + 6)y = 0$
b $y = a2^x + bx2^x$ $(E^2 - 4E + 4)y = 0$
c $y = a3^x + b(-2)^x$ $(\Delta^2 + \Delta - 6)y = 0$

6 Find the sum of the cubes of the first n integers

7 Show that $\Delta(f_n g_n) = f_{n+1}\Delta g_n + g_n\Delta f_n = g_{n+1}\Delta f_n + f_n\Delta g_n$

8 If $h = 1$, show that $\Delta \sin ax = 2 \sin (a/2) \cos a(x + \tfrac{1}{2})$ and that

$$\Delta \cos ax = -2 \sin \frac{a}{2} \sin a(x + \tfrac{1}{2})$$

9 Express each of the following in terms of factorial "polynomials" of the type $(x)^{-(k)}$

a $\dfrac{1}{(x+2)(x+3)}$ **b** $\dfrac{x}{(x+1)(x+2)}$ **c** $\dfrac{x-1}{(x+1)(x+3)}$

10 What is $\displaystyle\sum_{k=1}^{n} \frac{k}{(k+1)(k+2)(k+3)}$?

11 Show that $(x)^{(a)}(x)^{(b)} \neq (x)^{(a+b)}$, but that $(x+a)^{(a)}(x)^{(b)} = (x+a)^{(a+b)}$

12 **a** Show that $\displaystyle\sum_{k=1}^{n} y_k = \frac{E^n - 1}{E - 1} y_1$, and then, by putting $E = 1 + \Delta$, show that

$$\sum_{k=1}^{n} y_k = \left[n + \frac{n(n-1)}{2!}\Delta + \frac{n(n-1)(n-2)}{3!}\Delta^2 + \qquad \right] y_1$$

b Using the results of part **a**, evaluate $\displaystyle\sum_{k=1}^{n} k^2$

13 Show that $f(x_0,x_1) = f(x_0)/(x_0 - x_1) + f(x_1)/(x_1 - x_0)$ and that

$$f(x_0,x_1,x_2) = \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)} + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)} + \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_1)}$$

What is the generalization of these results to divided differences of higher order?

14 Show that

$$f(x_0,x_1) = \frac{\begin{vmatrix} f(x_0) & f(x_1) \\ 1 & 1 \end{vmatrix}}{\begin{vmatrix} x_0 & x_1 \\ 1 & 1 \end{vmatrix}}$$

and that

$$f(x_0,x_1,x_2) = \frac{\begin{vmatrix} f(x_0) & f(x_1) & f(x_2) \\ x_0 & x_1 & x_2 \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} x_0^2 & x_1^2 & x_2^2 \\ x_0 & x_1 & x_2 \\ 1 & 1 & 1 \end{vmatrix}}$$

What is the generalization of these results to divided differences of higher order?

15 If we define $f(x_0,x_1, \quad ,x_{n-1},x_n,x_n) = \lim\limits_{x \to x_n} f(x_0,x_1, \quad ,x_{n-1},x_n,x)$, show that

$$f(x_0,x_1, \quad ,x_{n-1},x_n,x_n) = \frac{df(x_0,x_1, \quad ,x_{n-1},x)}{dx}\bigg|_{x=x_n}$$

4.2

Interpolation formulas

One of the most important applications of finite differences is to the problem of interpolation In courses such as algebra and trigonometry, where tables of the elementary functions must occasionally be used, it is customary to obtain values between adjacent entries by the method of proportional parts or linear interpolation As is well known, this procedure amounts to replacing the arc of the tabulated function over one tabular interval by its chord and then reading the required functional value from the chord rather than from the arc itself (Fig 4 1a)

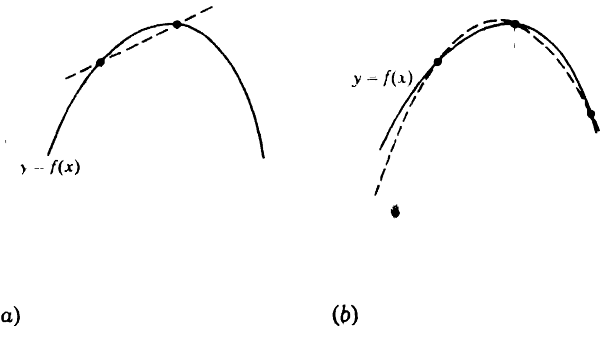


FIGURE 4 1 Straight-line and parabolic approximations to a given function

In this case the formula for the interpolated value turns out to be

(1) $$f(x_0 + rh) = f(x_0) + r[f(x_0 + h) - f(x_0)] = f_0 + r\,\Delta f_0$$

Obviously, if h is relatively large or if the graph of $f(x)$ is changing direction rapidly, the chord may not be a good approximation to the arc, and linear interpolation may involve a substantial error. One way to overcome this difficulty would be to approximate the graph of $f(x)$ by some curve which would "fit" the true arc more closely than a straight line could and then read the interpolated value from this approximating curve rather than from the chord (Fig. 4.1b). If, specifically, the graph of $f(x)$ is approximated over two successive tabular intervals by a parabola of the form $y = a + bx + cx^2$ chosen to pass through the three points

$$[x_0, f(x_0)] \qquad [x_0 + h, f(x_0 + h)] \qquad [x_0 + 2h, f(x_0 + 2h)]$$

the formula for the interpolated value is found without difficulty to be

$$\begin{aligned} f(x_0 + rh) &= f(x_0) + r[f(x_0 + h) - f(x_0)] \\ &\quad + \frac{r(r-1)}{2}[f(x_0 + 2h) - 2f(x_0 + h) + f(x_0)] \end{aligned}$$

(2) $$= f_0 + r\,\Delta f_0 + \frac{r(r-1)}{2!}\Delta^2 f_0$$

Proceeding in this fashion, using polynomial curves of higher and higher order to approximate the graph of $f(x)$, one could derive a succession of interpolation formulas involving higher and higher differences of the tabulated function and providing in general higher and higher accuracy in the interpolated values. In this section we shall obtain several important interpolation formulas, though we shall derive them by methods more general than the geometric approach we have just suggested.

Probably the most fundamental interpolation formula is **Newton's divided-difference formula:**

(3) $$\begin{aligned} f(x) = f(x_0) &+ (x - x_0)f(x_0,x_1) + (x - x_0)(x - x_1)f(x_0,x_1,x_2) + \cdots \\ &+ (x - x_0)(x - x_1)\cdots(x - x_{n-1})f(x_0,x_1,\ldots,x_n) \\ &+ (x - x_0)(x - x_1)\cdots(x - x_n)f(x,x_0,x_1,\ldots,x_n) \end{aligned}$$

From this all the other interpolation formulas of interest to us can easily be derived by suitably specializing the points x_0, x_1, . . . , x_n, which need not be regularly spaced or taken in consecutive order. For convenience in establishing (3) we shall restrict our discussion to some special, though adequately typical, value of n, say $n = 2$. Then, beginning with the third difference

$$f(x,x_0,x_1,x_2) = \frac{f(x,x_0,x_1) - f(x_0,x_1,x_2)}{x - x_2}$$

and solving for $f(x,x_0,x_1)$, we have

$$f(x,x_0,x_1) = f(x_0,x_1,x_2) + (x - x_2)f(x,x_0,x_1,x_2) \tag{4}$$

But $$f(x,x_0,x_1) = \frac{f(x,x_0) - f(x_0,x_1)}{x - x_1}$$

and, substituting this into (4) and solving for $f(x,x_0)$, we find

$$f(x,x_0) = f(x_0,x_1) + (x - x_1)f(x_0,x_1,x_2) + (x - x_1)(x - x_2)f(x,x_0,x_1,x_2) \tag{5}$$

Finally, since $$f(x,x_0) = \frac{f(x) - f(x_0)}{x - x_0}$$

we have, on substituting this into (5) and solving for $f(x)$,

$$\begin{aligned} f(x) = f(x_0) + (x - x_0)f(x_0,x_1) + (x - x_0)(x - x_1)f(x_0,x_1,x_2) \\ + (x - x_0)(x - x_1)(x - x_2)f(x,x_0,x_1,x_2) \end{aligned}$$

which is precisely Eq (3) in the special case $n = 2$ The extension of the preceding argument to any value of n is obvious

The last term in (3) differs from the other terms in that the divided difference appearing in it contains x as one of its arguments and, hence, is not to be found among the entries in the difference table of $f(x)$ For this reason the last term is usually referred to as the **remainder after** $n + 1$ **terms** or simply as the **error term,** and the interpolation series is often written in the form

$$f(x) = p_n(x) + r_{n+1}(x) \tag{6}$$

where, of course, $p_n(x)$ is the nth-degree polynomial

$$\begin{aligned} f(x_0) + (x - x_0)f(x_0,x_1) + (x - x_0)(x - x_1)f(x_0,x_1,x_2) + \cdots \\ + (x - x_0)(x - x_1) \cdots (x - x_{n-1})f(x_0,x_1, \ldots ,x_n) \end{aligned} \tag{7}$$

and $r_{n+1}(x)$ is the function

$$(x - x_0)(x - x_1) \cdots (x - x_n)f(x,x_0,x_1, \ldots ,x_n) \tag{8}$$

Using (6), (7), and (8), it is possible to obtain an interesting alternative expression for an nth divided difference and ultimately a somewhat more tractable form of the remainder term in (3) To do this, we observe that $r_{n+1}(x)$ vanishes at least $n + 1$ times on the closed interval between the largest and smallest values of the set $(x_0,x_1, \ldots ,x_n)$, since, in fact, it vanishes when $x = x_0, x_1, \ldots , x_n$ Therefore, assuming that the necessary derivatives exist, it follows from Rolle's theorem that $r'_{n+1}(x)$ must vanish at least n times on this interval, $r''_{n+1}(x)$ must vanish at least $n - 1$ times on this interval, and, continuing in this fashion, $r^{(n)}_{n+1}(x)$ must vanish at least once on this interval That is, there must exist at least one value of x, say $x = \xi$, between the largest and smallest values of the set $(x_0,x_1, \ldots ,x_n)$, such that $r^{(n)}_{n+1}(\xi) = 0$ Hence, differentiating (6) n times and evaluating the result for $x = \xi$, we have

$$f^{(n)}(\xi) - p_n^{(n)}(\xi) = r^{(n)}_{n+1}(\xi) = 0 \tag{9}$$

Now, from (7), the leading coefficient in the nth-degree polynomial $p_n(x)$ is $f(x_0,x_1,\ldots,x_n)$. Therefore,

$$p_n^{(n)}(\xi) = n!f(x_0,x_1,\ldots,x_n)$$

and, hence, from (9) we have, for each value of n,

$$f(x_0,x_1,\ldots,x_n) = \frac{f^{(n)}(\xi)}{n!} \tag{10}$$

where ξ is somewhere between the largest and smallest values of the set $(x_0,x_1,\ldots,x_n)$.

Applying (10) to the $(n+1)$st divided difference appearing in the expression for $r_{n+1}(x)$ in (8), we have, as an alternative form of $r_{n+1}(x)$,

$$r_{n+1}(x) = (x - x_0)(x - x_1)\cdots(x - x_n)\frac{f^{(n+1)}(\xi)}{(n+1)!} \tag{8a}$$

where now ξ is somewhere between the largest and smallest values of the set $(x,x_0,x_1,\ldots,x_n)$. The error term $r_{n+1}(x)$ is of great importance in theoretical studies of the convergence of the interpolation series (3), but the difficulty of estimating the factor

$$f(x,x_0,x_1,\ldots,x_n) = \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

often limits its usefulness in numerical work. Of course, if $f(x)$ is a polynomial of degree m, say, its divided differences of order greater than m are all exactly zero, and, if we extend the series (3) sufficiently far, the error term will be zero. In our work we shall neglect the error term in (3) on the assumption that, eventually, the divided differences become exactly zero or at least negligibly small, and that the series is extended to this point.

EXAMPLE 1

Find $f(2)$ from the following data.

x	$f(x)$	$f(x_i,x_j)$	$f(x_i,x_j,x_k)$	$f(x_i,x_j,x_k,x_l)$
−1.0	3.000			
		−5.000		
0.0	−2.000		5.500	
		3.250		−1.000
0.5	−0.375		3.500	
		6.750		−1.000
1.0	3.000		1.000	
		8.750		−1.000
2.5	16.125		−1.500	
		5.750		
3.0	19.000			

The construction of the difference table presents no problem, and, using Newton's formula, with $x_0 = 0$, we can write at once

$$\begin{aligned} f(2) &= -2.000 + (2-0)(3.250) + (2-0)(2-0.5)(3.500) \\ &\qquad + (2-0)(2-0.5)(2-1)(-1.000) \\ &= 12.000 \end{aligned}$$

In passing, we note that the ordinary process of linear interpolation yields the value $f(2) = 13.750$.

Closely associated with Newton's divided-difference formula is **Lagrange's interpolation formula**,*

(11)
$$
\begin{aligned}
f(x) = {} & \frac{(x - x_1)(x - x_2) \cdots (x - x_n)}{(x_0 - x_1)(x_0 - x_2) \cdots (x_0 - x_n)} f(x_0) \\
& + \frac{(x - x_0)(x - x_2) \cdots (x - x_n)}{(x_1 - x_0)(x_1 - x_2) \cdots (x_1 - x_n)} f(x_1) \\
& + \cdots \\
& + \frac{(x - x_0)(x - x_1) \cdots (x - x_{n-1})}{(x_n - x_0)(x_n - x_1) \cdots (x_n - x_{n-1})} f(x_n)
\end{aligned}
$$

Like Newton's divided-difference formula, this formula provides the equation of a polynomial of degree n (or less) which takes on $n + 1$ prescribed functional values when x takes on the values $x_0, x_1, \ldots, x_n$. Equation (11) can easily be derived from Eq. (3), but it is simpler merely to verify its properties. Clearly, it is a polynomial of degree n (or less), since each term on the right is a polynomial of degree n. Moreover, when $x = x_0$, every fraction except the first vanishes because of the factor $x - x_0$, and at the same time the first fraction reduces to 1, leaving just $f(x) = f(x_0)$, as required, when $x = x_0$. In the same way, when $x = x_1$, every fraction except the second becomes zero, and we have $f(x) = f(x_1)$. Similarly, we can verify without difficulty that $f(x)$ reduces to $f(x_2), f(x_3), \ldots, f(x_n)$ when $x = x_2, x_3, \ldots, x_n$, as required.

When the points $x_0, x_1, \ldots, x_n$ on which Newton's divided-difference formula is based are regularly spaced with tabular interval h, say, it is generally more convenient to express Formula (3) in terms of ordinary differences. To do this we observe that if

$$x = x_0 + rh \qquad \text{and} \qquad x_k = x_0 + kh$$

then $\qquad x - x_k = h(r - k) \qquad k = 0, 1, 2, \ldots, n$

and

(12)
$$(x - x_0)(x - x_1) \cdots (x - x_j) = h^{j+1} r(r - 1) \cdots (r - j)$$

Also, from Eq. (6), Sec. 4.1, we have

(13)
$$f(x_0, x_1, \ldots, x_{j+1}) = \frac{\Delta^{j+1} f_0}{(j + 1)!\, h^{j+1}}$$

Hence, substituting from (12) and (13) into (3), we find

(14)
$$
\begin{aligned}
f(x) & \equiv f(x_0 + rh) \\
& = f_0 + r\,\Delta f_0 + \frac{r(r - 1)}{2!} \Delta^2 f_0 + \frac{r(r - 1)(r - 2)}{3!} \Delta^3 f_0 + \cdots
\end{aligned}
$$

which is known as the **forward Gregory-Newton interpolation formula.**† Obviously this is a direct generalization of the formulas of linear and parabolic interpolation [Eqs. (1) and (2)]. Of course, the error term in (3) can be transformed into a corresponding error term for the series (14), but we shall leave this as an exercise.

---

* Named for the French mathematician Joseph Louis Lagrange (1736–1813).

† Co-named for the Scottish mathematician James Gregory (1661–1708).

For tables of limited extent, Formula (14) is especially adapted to interpolation near the upper end, i.e., for smaller values of x, and cannot conveniently be used near the lower end. For the latter case it would be desirable to have a formula using differences located above rather than below the point of interpolation. Such a formula can easily be derived by choosing the points $x_0, x_1, \ldots, x_n$ used in the divided-difference formula (3) to be the points

$$x_0, \quad x_0 - h, \quad x_0 - 2h, \quad \ldots, \quad x_0 - nh$$

Then $\quad x - x_k = h(r + k) \qquad k = 0, 1, 2, \ldots, n$

and

(15) $$(x - x_0)(x - x_1)\cdots(x - x_j) = h^{j+1}r(r + 1)\cdots(r + j)$$

Moreover, in this case the typical difference

$$f(x_0,x_1,\ldots,x_{j+1})$$

becomes $\quad f(x_0,x_{-1},\ldots,x_{-j-1})$

and, from the symmetry of divided differences, the last expression is equal to

$$f(x_{-j-1},x_{-j},\ldots,x_0)$$

Hence, using Eq. (7), Sec. 4.1 (with $n = k = j + 1$), we have, for our current choice of points,

(16) $$f(x_0,x_1,\ldots,x_{j+1}) = \frac{\Delta^{j+1}f_{-j-1}}{(j + 1)!h^{j+1}}$$

Hence, substituting from (15) and (16) into (3), we find

(17) $$\begin{aligned} f(x) &\equiv f(x_0 + rh) \\ &= f_0 + r\,\Delta f_{-1} + \frac{r(r + 1)}{2!}\Delta^2 f_{-2} \\ &\qquad + \frac{r(r + 1)(r + 2)}{3!}\Delta^3 f_{-3} + \cdots \end{aligned}$$

which is known as the **backward Gregory-Newton interpolation formula.**

EXAMPLE 2

Compute $f(1.03)$ from the following data

x	$f(x)$
1.00	1.000000
1.05	1.257625
1.10	1.531000
1.15	1.820875
1.20	2.128000

Δ	Δ^2	Δ^3
0.257625		
	0.015750	
0.273375		0.000750
	0.016500	
0.289875		0.000750
	0.017250	
0.307125		

The construction of the difference table presents no difficulty, and we need merely identify $x_0 = 1.00$, $h = 0.05$, $r = 0.6$ and then substitute into Formula (14)

$$
\begin{aligned}
f(1.03) &= f[1.00 + (0.6)(0.05)] \\
&= 1.000000 + (0.6)(0.257625) + \frac{(0.6)(0.6-1)}{2!}(0.015750) \\
&\qquad + \frac{(0.6)(0.6-1)(0.6-2)}{3!}(0.000750) \\
&= 1.152727
\end{aligned}
$$

Linear interpolation uses only the first two terms of the last series and hence yields the (presumably) less accurate value $f(1.03) = 1.154575$

There are various ways of obtaining central-difference interpolation formulas For instance, we can choose the points used in Newton's divided-difference formula in the following order

$$x_0 = x_0 \qquad x_1 = x_0 + h \qquad x_2 = x_0 - h \qquad x_3 = x_0 + 2h \qquad x_4 = x_0 - 2h, \qquad \ldots$$

Then substituting into (3) and using Eq (7), Sec 4.1, to simplify the various divided differences, we find

$$
\begin{aligned}
(18)\qquad f(x) &= f(x_0 + rh) \\
&= f_0 + r\,\Delta f_0 + \frac{r(r-1)}{2!}\Delta^2 f_{-1} + \frac{r(r-1)(r+1)}{3!}\Delta^3 f_{-1} \\
&\qquad + \frac{r(r-1)(r+1)(r-2)}{4!}\Delta^4 f_{-2} + \cdots
\end{aligned}
$$

or, introducing the central-difference operator δ by means of the operational equivalence $\Delta = \delta E^{1/2}$ [Eq (12), Sec 4.1],

$$
\begin{aligned}
(18a)\qquad f(x_0 + rh) &= f_0 + r\,\delta f_{1/2} + \frac{r(r-1)}{2!}\delta^2 f_0 + \frac{(r+1)r(r-1)}{3!}\delta^3 f_{1/2} \\
&\qquad + \frac{(r+1)r(r-1)(r-2)}{4!}\delta^4 f_0 + \cdots
\end{aligned}
$$

This is known as the **forward Newton-Gauss interpolation formula.**

In exactly the same way, by choosing the points $x_0, x_1, x_2, \ldots$ in the order

$$x_0 = x_0 \qquad x_1 = x_0 - h \qquad x_2 = x_0 + h \qquad x_3 = x_0 - 2h \qquad x_4 = x_0 + 2h, \qquad \ldots$$

and again substituting into (3) we obtain, after introducing the central-difference notation,

$$
\begin{aligned}
(19)\qquad f(x_0 + rh) &= f_0 + r\,\delta f_{-1/2} + \frac{(r+1)r}{2!}\delta^2 f_0 + \frac{(r+1)r(r-1)}{3!}\delta^3 f_{-1/2} \\
&\qquad + \frac{(r+2)(r+1)r(r-1)}{4!}\delta^4 f_0 + \cdots
\end{aligned}
$$

which is usually referred to as the **backward Newton-Gauss interpolation formula.**

If we take the average of Eqs (18a) and (19) we obtain a useful result known as **Stirling's interpolation formula:***

$$f(x_0 + rh) = f_0 + \frac{r}{1!}\frac{(\delta f_{1/2} + \delta f_{-1/2})}{2} + \frac{r^2}{2!}\delta^2 f_0 + \frac{r(r^2-1)}{3!}\frac{(\delta^3 f_{1/2} + \delta^3 f_{-1/2})}{2} + \frac{r^2(r^2-1)}{4!}\delta^4 f_0 + \cdots \tag{20}$$

Another formula of considerable utility can be obtained by eliminating the differences of odd order from Eq (18a) by means of the formulas $\delta f_{1/2} = f_1 - f_0$, $\delta^3 f_{1/2} = \delta^2 f_1 - \delta^2 f_0$, . . . This gives

$$\begin{aligned} f(x_0 + rh) = f_0 &+ r(f_1 - f_0) + \frac{r(r-1)}{2!}\delta^2 f_0 \\ &+ \frac{(r+1)r(r-1)}{3!}(\delta^2 f_1 - \delta^2 f_0) + \frac{(r+1)r(r-1)(r-2)}{4!}\delta^4 f_0 \\ &+ \frac{(r+2)(r+1)r(r-1)(r-2)}{5!}(\delta^4 f_1 - \delta^4 f_0) + \cdots \end{aligned}$$

or, collecting terms,

$$\begin{aligned} f(x_0 + rh) = -(r-1)f_0 &- \frac{r(r-1)(r-2)}{3!}\delta^2 f_0 \\ &- \frac{(r+1)r(r-1)(r-2)(r-3)}{5!}\delta^4 f_0 - \cdots \\ + r f_1 &+ \frac{(r+1)r(r-1)}{3!}\delta^2 f_1 \\ &+ \frac{(r+2)(r+1)r(r-1)(r-2)}{5!}\delta^4 f_1 + \cdots \end{aligned}$$

Finally, if we set $1 - r = s$ in the coefficients of the differences of f_0, we obtain the symmetric form

$$\begin{aligned} f(x_0 + rh) = s f_0 &+ \frac{s(s^2-1)}{3!}\delta^2 f_0 + \frac{s(s^2-1)(s^2-4)}{5!}\delta^4 f_0 + \cdots \\ + r f_1 &+ \frac{r(r^2-1)}{3!}\delta^2 f_1 + \frac{r(r^2-1)(r^2-4)}{5!}\delta^4 f_1 + \cdots \end{aligned} \tag{21}$$

which is known as the **Laplace-Everett interpolation formula.**

EXERCISES

1 Establish Eq (2) by finding the equation of the approximating parabola and evaluating it at $x = x_0 + rh$ (Hint Take x_0, x_1, x_2 to be 0, h, $2h$, respectively)

---

* Named for the Scottish mathematician James Stirling (1692–1770)

2 Obtain the error terms in the forward and backward Gregory-Newton formulas from the error term in Newton's divided-difference formula

3 Compute (a) $f(1.3)$ and (b) $f(1.95)$ from the following data

x	1.1	1.2	1.5	1.7	1.8	2.0
$f(x)$	1.112	1.219	1.636	2.054	2.323	3.011

4 Compute (a) $\sqrt{50.2}$ and (b) $\sqrt{55.9}$ from the following data

x	$\sqrt{x}$
50	7.07107
51	7.14143
52	7.21110
53	7.28011
54	7.34847
55	7.41620
56	7.48331

5 Fit a polynomial of minimum degree to the data of Example 1

6 Fit a polynomial of minimum degree to the following data

x	-1	1	2	4	5
$f(x)$	13	15	13	33	67

7 If y_0, y_1, y_2, y_3 are the values of a function at the equally spaced values x_0, x_1, x_2, x_3, show that the best estimate of the value of y corresponding to the value of x midway between x_1 and x_2 is

$$\frac{y_1 + y_2}{2} + \frac{(y_1 + y_2) - (y_0 + y_3)}{16}$$

8 Three readings are taken at equally spaced points $x = 0, h, 2h$ near the maximum (minimum) of a function $y = f(x)$ Show that the abscissa of the maximum (minimum) is approximately

$$\left(\frac{1}{2} - \frac{\Delta y_0}{\Delta^2 y_0}\right)h$$

and that the maximum (minimum) ordinate is approximately

$$y_1 - \frac{(\Delta y_1 + \Delta y_0)^2}{8\,\Delta^2 y_0}$$

9 Work Example 8, given that the three points where readings are taken are not equally spaced

10 Derive Lagrange's interpolation formula by expanding

$$\frac{f(x)}{(x - x_0)(x - x_1) \cdots (x - x_n)}$$

into partial fractions

4.3

Numerical integration and differentiation

Any of the interpolation formulas we obtained in the last section can be used to find the derivative of a tabular function. For instance, if we consider the forward Gregory-Newton formula

$$f(x_0 + rh) = f_0 + r\,\Delta f_0 + \frac{r(r-1)}{2!}\,\Delta^2 f_0 + \frac{r(r-1)(r-2)}{3!}\,\Delta^3 f_0 + \frac{r(r-1)(r-2)(r-3)}{4!}\,\Delta^4 f_0 + \cdots$$

and differentiate with respect to r, we find

(1) $$hf'(x_0 + rh) = \Delta f_0 + \frac{2r-1}{2}\,\Delta^2 f_0 + \frac{3r^2 - 6r + 2}{6}\,\Delta^3 f_0 + \frac{2r^3 - 9r^2 + 11r - 3}{12}\,\Delta^4 f_0 + \cdots$$

(2) $$h^2 f''(x_0 + rh) = \Delta^2 f_0 + (r-1)\,\Delta^3 f_0 + \frac{6r^2 - 18r + 11}{12}\,\Delta^4 f_0 + \cdots$$

(3) $$h^3 f'''(x_0 + rh) = \Delta^3 f_0 + \frac{2r-3}{2}\,\Delta^4 f_0 + \cdots$$

(4) $$h^4 f^{IV}(x_0 + rh) = \Delta^4 f_0 + \cdots$$

Specifically, if we put $r = 0$, we find for the successive derivatives at the tabular point x_0

(5) $$f'(x_0) = \frac{1}{h}\left(\Delta f_0 - \frac{1}{2}\,\Delta^2 f_0 + \frac{1}{3}\,\Delta^3 f_0 - \frac{1}{4}\,\Delta^4 f_0 + \cdots\right)$$

(6) $$f''(x_0) = \frac{1}{h^2}\left(\Delta^2 f_0 - \Delta^3 f_0 + \frac{11}{12}\,\Delta^4 f_0 - \cdots\right)$$

(7) $$f'''(x_0) = \frac{1}{h^3}\left(\Delta^3 f_0 - \frac{3}{2}\,\Delta^4 f_0 + \cdots\right)$$

(8) $$f^{IV}(x_0) = \frac{1}{h^4}\left(\Delta^4 f_0 - \cdots\right)$$

Similarly, from the backward Gregory-Newton formula we obtain

(9) $$hf'(x_0 + rh) = \Delta f_{-1} + \frac{2r+1}{2}\,\Delta^2 f_{-2} + \frac{3r^2 + 6r + 2}{6}\,\Delta^3 f_{-3} + \frac{2r^3 + 9r^2 + 11r + 3}{12}\,\Delta^4 f_{-4} + \cdots$$

(10) $$h^2 f''(x_0 + rh) = \Delta^2 f_{-2} + (r+1)\,\Delta^3 f_{-3} + \frac{6r^2 + 18r + 11}{12}\,\Delta^4 f_{-4} + \cdots$$

(11) $$h^3 f'''(x_0 + rh) = \Delta^3 f_{-3} + \frac{2r+3}{2}\,\Delta^4 f_{-4} + \cdots$$

(12) $$h^4 f^{IV}(x_0 + rh) = \Delta^4 f_{-4} + \cdots$$

and, at the point x_0,

$$f'(x_0) = \frac{1}{h}\left(\Delta f_{-1} + \frac{1}{2}\Delta^2 f_{-2} + \frac{1}{3}\Delta^3 f_{-3} + \frac{1}{4}\Delta^4 f_{-4} + \quad \right) \tag{13}$$

$$f''(x_0) = \frac{1}{h^2}\left(\Delta^2 f_{-2} + \Delta^3 f_{-3} + \frac{11}{12}\Delta^4 f_{-4} + \quad \right) \tag{14}$$

$$f'''(x_0) = \frac{1}{h^3}\left(\Delta^3 f_{-3} + \frac{3}{2}\Delta^4 f_{-4} + \quad \right) \tag{15}$$

$$f^{IV}(x_0) = \frac{1}{h^4}\left(\Delta^4 f_{-4} + \quad \right) \tag{16}$$

For a rigorous development, an error term analogous to Eq (8a), Sec 4.2, should be found for any formula of numerical differentiation. This can be done, but the results are of relatively little use in routine calculations, and we shall not take them into account. However, it should be borne in mind that, unless we are dealing with a polynomial, numerical differentiation may involve errors of considerable magnitude, the errors increasing significantly as derivatives of higher order are computed.

EXAMPLE 1

Find the first and second derivatives of $\sqrt{x}$ at $x = 2.5$ from the table

x	$\sqrt{x}$	Δ	Δ^2
2.50	1.58114		
		0.01573	
2.55	1.59687		−0.00015
		0.01558	
2.60	1.61245		−0.00015
		0.01543	
2.65	1.62788		0.00014
		0.01529	
2.70	1.64317		−0.00015
		0.01514	
2.75	1.65831		

Using Eqs (5) and (6) with $x_0 = 2.50$ and $h = 0.05$, we find at once

$$f'(2.5) = \frac{1}{0.05}\left[0.01573 - \frac{1}{2}(-0.00015)\right] = 0.3160$$

$$f''(2.5) = \frac{1}{(0.05)^2}(-0.00015) = -0.0600$$

The correct values to four decimal places are, of course,

$$f'(2.5) = \left.\frac{1}{2\sqrt{x}}\right|_{x=2.5} = 0.3162$$

$$f''(2.5) = \left.\frac{-1}{4x\sqrt{x}}\right|_{x=2.5} = -0.0632$$

To obtain formulas for numerical integration, it is convenient to begin by considering the related problem of the summation of series, a topic on which we touched briefly at the end of Sec 4 1 In doing this it will be convenient to use certain additional operational equivalences which we shall now develop We begin with Maclaurin's expansion,

(17) $$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \frac{h^3}{3!}f'''(x) +$$

or, introducing the operators E and $D \equiv d/dx$,

(18) $$Ef(x) = \left(1 + hD + \frac{h^2D^2}{2!} + \frac{h^3D^3}{3!} + \quad\right)f(x)$$

Now, the series on the right is simply the expansion of the exponential e^{hD} Hence, we can write (18) in the form

$$Ef(x) = e^{hD}f(x)$$

from which we infer the operational equivalences

(19) $$E = e^{hD}$$

(20) $$\Delta \equiv E - 1 = e^{hD} - 1$$

Next, we introduce the integration operator

$$If(x) = \int_x^{x+h} f(x)\,dx$$

Then $$IDf(x) = \int_x^{x+h} f'(x)\,dx = f(x+h) - f(x) = \Delta f(x)$$

and, if $F(x)$ is any antiderivative of $f(x)$,

$$DIf(x) = D\int_x^{x+h} f(x)\,dx = D[F(x+h) - F(x)]$$
$$= f(x+h) - f(x) = \Delta f(x)$$

Hence, D and I commute with each other, and we have the further equivalences

(21) $$ID = DI = \Delta$$

We are now in a position to establish the famous **Euler-Maclaurin summation formula**:

(22) $$\sum_{i=0}^{n} f_i = \frac{1}{h}\int_{x_0}^{x_n} f(x)\,dx + \frac{1}{2}(f_0 + f_n) + \sum_{i=1}^{\infty} \frac{B_{2i}}{(2i)!}h^{2i-1}\left(f_n^{(2i-1)} - f_0^{(2i-1)}\right)$$

where the B's are the Bernoulli numbers $B_2 = \frac{1}{6}$, $B_4 = -\frac{1}{30}$, to be defined below. We begin by writing, with the aid of Eq (21),

$$h\,\Delta f(x) = hDIf(x)$$

or, replacing Δ by its equivalent from Eq (20),

$$h(e^{hD} - 1)f(x) = hDIf(x)$$

or further,

(23) $$hf(x) = \frac{hD}{e^{hD} - 1}\,If(x)$$

It is now necessary to expand the fractional operator $hD/(e^{hD}-1)$ in a power series in hD. This can be done in various ways, but perhaps the simplest is to replace e^{hD} by its series equivalent and then make use of the method of undetermined coefficients. Thus we have

$$\frac{hD}{\left(1+hD+\dfrac{h^2D^2}{2!}+\dfrac{h^3D^3}{3!}+\cdots\right)-1} = a_0 + a_1hD + \frac{a_2}{2!}h^2D^2 + \frac{a_3}{3!}h^3D^3 + \cdots$$

or, simplifying the fraction on the left and then clearing fractions,

$$1 = \left(1+\frac{hD}{2!}+\frac{h^2D^2}{3!}+\frac{h^3D^3}{4!}+\cdots\right)\left(a_0 + a_1hD + \frac{a_2}{2!}h^2D^2 + \frac{a_3}{3!}h^3D^3 + \cdots\right)$$

Now, multiplying the two series and equating the coefficients of like powers of hD on the two sides of this identity, we obtain the equations

$$a_0 = 1 \qquad \frac{a_0}{2!} + a_1 = 0 \qquad \frac{a_0}{3!} + \frac{a_1}{2!} + \frac{a_2}{2!} = 0$$

$$\frac{a_0}{4!} + \frac{a_1}{3!} + \frac{a_2}{2!2!} + \frac{a_3}{3!} = 0$$

$$\frac{a_0}{5!} + \frac{a_1}{4!} + \frac{a_2}{3!2!} + \frac{a_3}{2!3!} + \frac{a_4}{4!} = 0$$

from which we find without difficulty

$$a_0 = 1 \qquad a_1 = -\tfrac{1}{2} \qquad a_2 = \tfrac{1}{6} \qquad a_3 = 0 \qquad a_4 = -\tfrac{1}{30},$$

The function $e^x/(e^x-1)$ occurs in numerous applications, and the coefficients $\{a_i\}$ in its expansion have many interesting and important properties. These coefficients are ordinarily referred to as the **Bernoulli numbers** $\{B_i\}$, and formulas have been developed which give them explicitly for any value of i. For our purposes we need to know only the numerical values of the first few B's and the fact that after B_1 all B's with odd subscripts are zero. Thus we can write

$$\frac{hD}{e^{hD}-1} = \sum_{i=0}^{\infty}\frac{B_i}{i!}(hD)^i = 1 - \frac{hD}{2} + \frac{h^2D^2}{12} - \frac{h^4D^4}{720} + \cdots$$

and hence, returning to Eq. (23),

$$f(x) = \frac{1}{h}\left[\sum_{i=0}^{\infty}\frac{B_i}{i!}(hD)^i\right] If(x)$$

or, detaching the first term from the series and factoring hD from the remaining terms,

$$f(x) = \frac{1}{h}\left[1 + hD \sum_{i=1}^{\infty} \frac{B_i}{i!}(hD)^{i-1}\right] If(x)$$

$$= \frac{1}{h} If(x) + \left[\sum_{i=1}^{\infty} \frac{B_i}{i!}(hD)^{i-1}\right] DIf(x)$$

(24) $$= \frac{1}{h}\int_{x}^{x+h} f(x)\,dx + \left[\sum_{i=1}^{\infty} \frac{B_i}{i!}(hD)^{i-1}\right] \Delta f(x)$$

Now let us evaluate Eq. (24) for $x = x_0, x_1, \ldots, x_{n-1}$ and add the results, recalling that

$$\Delta f_0 + \Delta f_1 + \cdots + \Delta f_{n-1} = f_n - f_0$$

$$f_0 = \frac{1}{h}\int_{x_0}^{x_1} f(x)\,dx + \left[\sum_{i=1}^{\infty} \frac{B_i}{i!}(hD)^{i-1}\right] \Delta f_0$$

$$f_1 = \frac{1}{h}\int_{x_1}^{x_2} f(x)\,dx + \left[\sum_{i=1}^{\infty} \frac{B_i}{i!}(hD)^{i-1}\right] \Delta f_1$$

$$f_{n-1} = \frac{1}{h}\int_{x_{n-1}}^{x_n} f(x)\,dx + \left[\sum_{i=1}^{\infty} \frac{B_i}{i!}(hD)^{i-1}\right] \Delta f_{n-1}$$

$$\sum_{i=0}^{n-1} f_i = \frac{1}{h}\int_{x_0}^{x_n} f(x)\,dx + \left[\sum_{i=1}^{\infty} \frac{B_i}{i!}(hD)^{i-1}\right] (f_n - f_0)$$

Since $B_1 = -\frac{1}{2}$ and $B_3 = B_5 = B_7 = \cdots = 0$, the last formula can be simplified somewhat by detaching the first term from the sum on the right-hand side and then setting $i = 2j$ in the rest of the series

$$\sum_{i=0}^{n-1} f_i = \frac{1}{h}\int_{x_0}^{x_n} f(x)\,dx - \frac{1}{2}(f_n - f_0) + \sum_{j=1}^{\infty} \frac{B_{2j}}{(2j)!} h^{2j-1}\left(f_n^{(2j-1)} - f_0^{(2j-1)}\right)$$

Finally, if we add f_n to both members of this identity we obtain Formula (22), as required.

If Eq. (22) is solved for the integral, we obtain

(25) $$\int_{x_0}^{x_n} f(x)\,dx = h\sum_{i=0}^{n} f_i - \frac{h}{2}(f_0 + f_n) - \sum_{j=1}^{\infty} \frac{B_{2j}}{(2j)!} h^{2j}\left(f_n^{(2j-1)} - f_0^{(2j-1)}\right)$$

which is a fundamental formula of numerical integration. Equation (25) is especially adapted to the integration of functions defined by analytic expressions which can conveniently be differentiated. For functions defined only by a table of values it is usually more convenient to have an integration formula in which the "correction terms" are expressed as differences rather than as derivatives. To obtain such a formula from Eq. (25), we need only replace the derivatives $f_0', f_0''', \ldots$ by means of Eqs. (5), (7), . . . and the derivatives $f_n', f_n''', \ldots$ by means of Eqs.

(13), (15), . This gives us

$$\int_{x_0}^{x_n} f(x)\,dx = h\left(\frac{f_0}{2} + f_1 + \qquad + f_{n-1} + \frac{f_n}{2}\right)$$

$$- \frac{h^2}{12}\left[\frac{1}{h}\left(\Delta f_{n-1} + \frac{\Delta^2 f_{n-2}}{2} + \frac{\Delta^3 f_{n-3}}{3} + \frac{\Delta^4 f_{n-4}}{4} + \qquad\right)\right.$$

$$\left. - \frac{1}{h}\left(\Delta f_0 - \frac{\Delta^2 f_0}{2} + \frac{\Delta^3 f_0}{3} - \frac{\Delta^4 f_0}{4} + \qquad\right)\right]$$

$$+ \frac{h^4}{720}\left[\frac{1}{h^3}\left(\Delta^3 f_{n-3} + \frac{3}{2}\Delta^4 f_{n-4} + \qquad\right)\right.$$

$$\left. - \frac{1}{h^3}\left(\Delta^3 f_0 - \frac{3}{2}\Delta^4 f_0 + \cdot \qquad\right)\right]$$

$$+$$

$$= h\left(\frac{f_0}{2} + f_1 + \qquad + f_{n-1} + \frac{f_n}{2}\right) - \frac{h}{12}(\Delta f_{n-1} - \Delta f_0) \tag{26}$$

$$- \frac{h}{24}(\Delta^2 f_{n-2} + \Delta^2 f_0) - \frac{19h}{720}(\Delta^3 f_{n-3} - \Delta^3 f_0)$$

$$- \frac{3h}{160}(\Delta^4 f_{n-4} + \Delta^4 f_0) -$$

which is known as **Gregory's formula of numerical integration.** In passing, we note that both (25) and (26) reduce to the well-known trapezoidal rule of integration if the correction terms are neglected.

EXAMPLE 2

Compute $\int_0^1 f(x)\,dx$ for the function defined by the following table

x	$f(x)$	Δ	Δ^2	Δ^3	Δ^4
0.0	0.4698220				
		0.0144778			
0.2	0.4842998		−0.0004670		
		0.0140108		0.0000290	
0.4	0.4983106		−0.0004380		−0.0000024
		0.0135728		0.0000266	
0.6	0.5118834		−0.0004114		−0.0000023
		0.0131614		0.0000243	
0.8	0.5250448		−0.0003871		
		0.0127743			
1.0	0.5378191				

Using Eq (26), we have at once

$$\int_0^1 f(x)\,dx = \frac{1}{5}\left(\frac{0.4698220}{2} + 0.4842998 + 0.4983106 + 0.5118834 + 0.5250448 + \frac{0.5378191}{2}\right)$$

$$- \frac{1}{60}(0.0127743 - 0.0144778) - \frac{1}{120}(-0.0003871 - 0.0004670)$$

$$- \frac{19}{3{,}600}(0.0000243 - 0.0000290) -$$

$$= 0.5047073$$

The integral in this problem is actually $\int_0^1 \log(2.95 + 0.5x)\,dx$, and its exact value is easily found to be 0.5047074, correct to seven decimal places. The approximate value is, therefore, in error by only 0.0000001.

In many important applications it is necessary to compute a **running integral** of a tabular function, i.e., an integral of the form

$$\int_{x_0}^{x} f(x)\,dx$$

where x takes on successively each of the values at which $f(x)$ is tabulated. For such a calculation the familiar **trapezoidal rule**:

$$\int_a^b f(x)\,dx = h\left(\frac{f_0}{2} + f_1 + f_2 + \cdots + f_{n-2} + f_{n-1} + \frac{f_n}{2}\right) \tag{27}$$

is especially well adapted. For if to the given table we adjoin a column of the averages

$$\frac{f_0 + f_1}{2}, \quad \frac{f_1 + f_2}{2}, \quad \ldots$$

the required integrals are precisely the sums of the entries in this column from the top down to each entry in turn, multiplied by h. Moreover, each sum can be found from the preceding one by adding to it the next average. Table 4.1 shows the computational pattern in detail.

By recording each average in the cell above the one where it appears in this table and then summing the column of averages from the bottom upward, the process can also be adapted to the calculation of running integrals of the form

$$\int_{x}^{x_n} f(x)\,dx$$

table 4.1

x	$f(x)$	Average	$\sum$	$h\sum = \int_{x_0}^{x} f(x)\,dx$
x_0	f_0	—	$\sum_0 = 0$	$h\sum_0 = \int_{x_0}^{x_0} f(x)\,dx$
x_1	f_1	$\frac{f_0+f_1}{2}$	$\sum_1 = \sum_0 + \frac{f_0+f_1}{2} = \frac{f_0}{2} + \frac{f_1}{2}$	$h\sum_1 = \int_{x_0}^{x_1} f(x)\,dx$
x_2	f_2	$\frac{f_1+f_2}{2}$	$\sum_2 = \sum_1 + \frac{f_1+f_2}{2} = \frac{f_0}{2} + f_1 + \frac{f_2}{2}$	$h\sum_2 = \int_{x_0}^{x_2} f(x)\,dx$
x_3	f_3	$\frac{f_2+f_3}{2}$	$\sum_3 = \sum_2 + \frac{f_2+f_3}{2} = \frac{f_0}{2} + f_1 + f_2 + \frac{f_3}{2}$	$h\sum_3 = \int_{x_0}^{x_3} f(x)\,dx$
			.	

EXAMPLE 3

Compute $\int_0^x e^x\,dx$ for $x = 0.2, 0.4, 0.6, \ldots, 1.8, 2.0$ from the following table.

x	e^x	Average	$\sum$	$h\sum = \int_0^x e^x\,dx$	True value $= e^x - 1$
0.0	1.000		0.000	0.000	0.000
0.2	1.221	1.110	1.110	0.222	0.221
0.4	1.492	1.356	2.466	0.493	0.492
0.6	1.822	1.657	4.123	0.825	0.822
0.8	2.226	2.024	6.147	1.229	1.226
1.0	2.718	2.472	8.619	1.724	1.718
1.2	3.320	3.019	11.638	2.328	2.320
1.4	4.055	3.687	15.325	3.065	3.055
1.6	4.953	4.504	19.829	3.966	3.953
1.8	6.050	5.501	25.330	5.066	5.050
2.0	7.389	6.719	32.049	6.410	6.389

Following the computational scheme described above, with $h = 0.2$, we obtain the required integrals without difficulty. For purposes of comparison, the theoretical values of the integral, correct to three decimal places are also shown. Had the interval between successive entries been smaller, say $h = 0.1$, the accuracy would have been significantly higher, although for $h = 0.2$ the last, and, therefore, the least accurate integral is in error by only

$$\frac{6.410 - 6.389}{6.389} \times 100 = 0.3\%$$

EXERCISES

1. From the data in the following table compute the first three derivatives of $\ln x$ at $x = 200$ and at $x = 205$, and check these against the exact values.

x	$\ln x$
200	5.29831737
201	5.30330491
202	5.30826770
203	5.31320598
204	5.31811999
205	5.32300998

2. Using the data of Exercise 1, compute the first three derivatives of $\ln x$ at $x = 200.5$ and at $x = 204.5$, and check these against the exact values.
3. Use Eq. (22) to obtain a formula for the sum of the cubes of the first n integers.
4. Using Eq. (22), establish the relation

$$\sum_{i=a}^{\infty} \frac{1}{i^2} \sim \frac{1}{a} + \frac{1}{2a^2} + \frac{1}{6a^3} - \frac{1}{30a^5} + \cdots + \frac{B_{2n}}{a^{2n+1}} + \cdots$$

(Note that the equality sign is not used between the two members of the given relation. In fact, although B_{2n} at first decreases with n, it increases more rapidly than a^{2n+1} for any fixed a as $n \to \infty$; hence the series on the right diverges and certainly cannot be said to equal the sum on the left. Interestingly enough, however, the series is a good approximation to the sum if it is terminated at the point where its terms are a minimum. Divergent series with this property are usually referred to as **asymptotic series.**)

5 What is $\sum_{i=1}^{1,000} \frac{1}{i}$?

6 Compute $\int_0^x e^{-x^2}\,dx$ for values of x at intervals of $h = 0.1$ from 0 to 1

7 Compute $\int_x^1 \frac{\sin x}{x}\,dx$ for values of x at intervals of $h = 0.1$ from 0 to 1.

8 Compute $\int_0^1 \sin x^2\,dx$ by the trapezoidal rule with $h = 0.1$, and compare your answer with the results given by Eq. (25)

a when only the first-derivative correction terms are included, and

b when the first- and third-derivative correction terms are included

9 Work Exercise 8 using Eq. (26) and the corresponding difference correction terms

10 Show that, when $n = 2$ and differences of order higher than the second are neglected, Eq. (26) reduces to

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{h}{3}(f_0 + 4f_1 + f_2)$$

By applying this to successive pairs of tabular intervals, establish **Simpson's rule**:

$$\int_{x_0}^{x_n} f(x)\,dx = \frac{h}{3}(f_0 + 4f_1 + 2f_2 + \cdots + 2f_{n-2} + 4f_{n-1} + f_n) \qquad n \text{ even}$$

11 By making two applications of the formula of the last exercise, establish the following formula for the numerical evaluation of double integrals

$$\int_{y_0}^{y_2} \int_{x_0}^{x_2} f(x,y)\,dx\,dy = \frac{hk}{9}\,[(f_{00} + f_{02} + f_{20} + f_{22}) + 4(f_{01} + f_{10} + f_{12} + f_{21}) + 16f_{11}]$$

where h and k are, respectively, the intervals at which x and y are tabulated, and $f_{ij} = f(x_0 + ih, y_0 + jk)$. Give the generalization of this result to integrals of the form

$$\int_{y_0}^{y_{2n}} \int_{x_0}^{x_{2m}} f(x,y)\,dx\,dy$$

12 Establish the operational equivalence $D = (1/h)\ln(1 + \Delta)$, and then, by expanding the right member into an infinite series, establish Formulas (5) to (8). [Hint. Begin with Eq. (20).]

13 Establish the operational equivalence $D = (2/h)\sinh^{-1}\delta/2$, and use it to develop a central difference formula for numerical differentiation. [Hint. Begin with Eq. (14), Sec. 4.1, and Eq. (19).]

14 If Lagrange's interpolation formula is integrated from x_0 to x_n, a formula of numerical integration of the form

$$\int_{x_0}^{x_n} f(x)\,dx = c_0f_0 + c_1f_1 + \cdots + c_{n-1}f_{n-1} + c_nf_n$$

is obtained. Determine the values of the c's when $n = 2$, and thence derive Simpson's rule (see Exercise 10).

15 Obtain the values of the c's in the formula of Exercise 14

a When $n = 3$ **b** When $n = 4$

4.4

The numerical solution of differential equations

One of the most important applications of finite differences is to the numerical solution of differential equations which, because of their complexity, cannot be solved by exact methods. Many procedures are available for doing this,* some of considerable generality, others especially adapted to equations of a particular form. Of the many methods which have been devised we shall present only the *method of Milne* and the *Runge-Kutta method*. These can be applied to simultaneous differential equations as well as to single equations of any order and are, therefore, adequate for almost any problem one is likely to encounter.

The fundamental problem is to find the solution of the first-order differential equation

$$\frac{dy}{dx} = f(x, y) \tag{1}$$

which satisfies the initial condition $y = y_0$ when $x = x_0$. We do not, of course, expect to find an equation for the solution. Instead, our object is merely to plot or tabulate the solution curve point by point, beginning at (x_0, y_0) and continuing thereafter at selected values of x, usually equally spaced, until the solution has been extended over the required range.

To develop **Milne's method** we begin with Eq. (1), Sec. 4.3, written in terms of y rather than f, and evaluate it for $r = 1, 2, 3,$ and 4, getting

$$y_1' = \frac{1}{h}\left(\Delta y_0 + \frac{1}{2}\Delta^2 y_0 - \frac{1}{6}\Delta^3 y_0 + \frac{1}{12}\Delta^4 y_0 + \qquad\right)$$

$$y_2' = \frac{1}{h}\left(\Delta y_0 + \frac{3}{2}\Delta^2 y_0 + \frac{1}{3}\Delta^3 y_0 - \frac{1}{12}\Delta^4 y_0 + \qquad\right)$$

$$y_3' = \frac{1}{h}\left(\Delta y_0 + \frac{5}{2}\Delta^2 y_0 + \frac{11}{6}\Delta^3 y_0 + \frac{1}{4}\Delta^4 y_0 + \qquad\right)$$

$$y_4' = \frac{1}{h}\left(\Delta y_0 + \frac{7}{2}\Delta^2 y_0 + \frac{13}{3}\Delta^3 y_0 + \frac{25}{12}\Delta^4 y_0 + \qquad\right)$$

or, neglecting differences beyond the fourth and replacing the remaining differences by their equivalent expressions in terms of the successive functional values [Eqs (15*a*), Sec. 4 1],

---

* See, for instance, H. Levy and E. A. Baggott, "Numerical Studies in Differential Equations," vol. 1, C. A. Watts & Co., Ltd., London, 1934, and W. E. Milne, "Numerical Solutions of Differential Equations," John Wiley & Sons, Inc., New York, 1953.

$$
\begin{aligned}
y_1' &= \frac{1}{12h}(-3y_0 - 10y_1 + 18y_2 - 6y_3 + y_4) \\
y_2' &= \frac{1}{12h}(y_0 - 8y_1 + 8y_3 - y_4) \\
y_3' &= \frac{1}{12h}(-y_0 + 6y_1 - 18y_2 + 10y_3 + 3y_4) \\
y_4' &= \frac{1}{12h}(3y_0 - 16y_1 + 36y_2 - 48y_3 + 25y_4)
\end{aligned}
\tag{2}
$$

Now, if we subtract the second equation in the last set from twice the sum of the first and third equations and solve the result for y_4, we obtain

$$y_4 = y_0 + \frac{4h}{3}(2y_1' - y_2' + 2y_3')$$

or, in more general terms,

$$y_{n+1} = y_{n-3} + \frac{4h}{3}(2y_{n-2}' - y_{n-1}' + 2y_n') \tag{3}$$

If we know the values of y and y' down to and including their values at x_n, Eq (3) thus enables us to "reach out" one step further and compute y_{n+1} With y_{n+1} known, we can then return to the given differential equation (1) and compute y_{n+1}' Then using Eq (3) again, we can find y_{n+2}, and so on, step by step, until the solution has been extended over the desired range All that remains is to devise a means of finding enough y's and y''s to get the process under way

One possibility is to begin the tabulation of y by expanding it in a Taylor series around the point $x = x_0$

$$y = y_0 + y_0'(x - x_0) + y_0''\frac{(x - x_0)^2}{2!} + y_0'''\frac{(x - x_0)^3}{3!} + \cdot \tag{4}$$

The value of y_0 is, of course, given The value of y_0' can be found at once by substituting x_0 and y_0 into the given differential equation (1) To find the second derivative we need only differentiate the given equation, getting

$$y'' = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}y' \tag{5}$$

Since $f(x,y)$ is a given function, its partial derivatives are known and become definite numbers when x_0 and y_0 are substituted into them. Moreover, the value of y' at (x_0,y_0) has already been found, and thus (5) furnishes the value of y_0''. Similarly, differentiating (5) and evaluating the result at (x_0,y_0) will give y_0''', and so on In this way the first few terms of the expansion of y around the point (x_0,y_0) can be constructed In especially favorable cases the general term of the series (4) can be found and the region of convergence established When this happens, (4) is the required solution, and we need look no further In general, however, successive differentiation of $f(x,y)$ becomes too complicated to continue or the resulting series converges too slowly to be of practical value, and we must fall back on Milne's or some similar method

With (4) available as a representation of y in the neighborhood of $x = x_0$, we can set $x = x_0 + h \equiv x_1$ and calculate y_1 Similarly, setting $x = x_0 + 2h$ and $x_0 + 3h$, we can find y_2 and y_3 Then, substituting (x_1,y_1), (x_2,y_2), and (x_3,y_3) into the given differential equation, we can compute y_1', y_2', and y_3' without difficulty With these values we are then in a position to begin the step-by-step solution of the differential equation by means of Eq. (3)

From the preceding discussion it is clear that Eq (3) is in general adequate for the step-by-step solution of $y' = f(x,y)$. However, as a precaution against errors of various kinds, it is desirable to have a second, independent formula into which y_{n+1} can be substituted as a check To obtain such an equation we return to (2) and add 4 times the third equation to the sum of the second and fourth and solve the resulting equation for y_4, getting

$$y_4 = y_2 + \frac{h}{3}(y_2' + 4y_3' + y_4')$$

or, in more general terms,

$$y_{n+1} = y_{n-1} + \frac{h}{3}(y_{n-1}' + 4y_n' + y_{n+1}') \tag{6}$$

This formula cannot be used as a formula of extrapolation, since it involves y_{n+1}', which cannot be found unless y_{n+1} is already known However, after y_{n+1} has been found by means of (3), y_{n+1}' can be calculated, and enough information is then available to permit the use of (6) If the value of y_{n+1}, as given by (6), agrees with the value found from (3), we are ready to move on to the calculation of y_{n+2} On the other hand, if the two values of y_{n+1} do not agree, we must use the second value of y_{n+1} to compute a new value of y_{n+1}', substitute these into (6), and continue the process until two successive values of y_{n+1} are in agreement When this happens, we are ready to continue the tabulation of y by returning to (3) and determining an initial estimate of y_{n+2}

Formulas like (3), which express a new value exclusively in terms of quantities already found, are known as **open formulas** or **predictor formulas.** Those, like (6), which express a new value in terms of one or more additional new quantities and which, therefore, can be used only for purposes of checking and refining are known as **closed formulas** or **corrector formulas.**

The method of Milne is readily extended to the solution of simultaneous and higher-order equations. For instance, if we have the two equations

$$y' = f(x,y,z) \qquad \text{and} \qquad z' = g(x,y,z)$$

with the initial conditions $y = y_0$, $z = z_0$ when $x = x_0$, and if by independent means we have calculated (y_1,y_2,y_3), (z_1,z_2,z_3), and the related quantities (y'_1,y'_2,y'_3) and (z'_1,z'_2,z'_3), then, using Eq (3) and an identical version with z replacing y, we can compute y_4 and z_4 After that, we can compute y'_4 and z'_4 from the differential equations and again use (3) to obtain y_5 and z_5, and so on as far as desired. Of course, the closed formula (6) can be used to check and correct both y_{n+1} and z_{n+1} if and when this is deemed necessary

The application of Milne's method to equations of higher order is now immediate, since such an equation can always be replaced by a system of simultaneous first-order equations For instance $y'' = g(x,y,y')$ is equivalent to the system

$$y' = z \qquad z' = g(x,y,z)$$

which is just a special case, with $f(x,y,z) \equiv z$, of the general problem of two simultaneous first-order equations

The Runge-Kutta method differs from Milne's method in several significant respects In the first place, it requires no predetermination of a set of starting values and, hence, is completely self-contained Second, it does not require the values of x at which the solution is being tabulated to be equally spaced; hence, the interval between successive values of x can be varied throughout the process, as time and accuracy may require

The Runge-Kutta method can be thought of as a generalization of the following extremely simple (and quite inaccurate) procedure If one is given the first-order differential equation

$$\frac{dy}{dx} = f(x,y)$$

and the initial condition $y = y_0$ when $x = x_0$, the value of y, say $y_1 = y_0 + \Delta y$, at $x_1 = x_0 + \Delta x$ can be approximated by using the usual differential estimate of the increment Δy,

$$\Delta y = \left.\frac{dy}{dx}\right|_{x_0,y_0} \Delta x = f(x_0,y_0)\,\Delta x$$

With this value for Δy available, an approximate value for $y_1 = y_0 + \Delta y$, namely,

$$y_1 = y_0 + \left.\frac{dy}{dx}\right|_{x_0,y_0} \Delta x = y_0 + f(x_0,y_0)\,\Delta x \tag{7}$$

is determined, and the process can be repeated to obtain y_2, y_3,

On the other hand, having a first approximation to y_1, one can compute dy/dx at the point (x_1,y_1) and then use the average value

$$\frac{1}{2}\left(\frac{dy}{dx}\bigg|_{x_0,y_0} + \frac{dy}{dx}\bigg|_{x_1,y_1}\right) \qquad \text{instead of} \qquad \frac{dy}{dx}\bigg|_{x_0,y_0}$$

in Eq. (7), to obtain a (presumably) more accurate value for y_1, namely,

(8) $$y_1 = y_0 + \frac{1}{2}\left(\frac{dy}{dx}\bigg|_{x_0,y_0} + \frac{dy}{dx}\bigg|_{x_1,y_1}\right)\Delta x = y_0 + \frac{f(x_0,y_0) + f(x_1,y_1)}{2}\,\Delta x$$

before attempting to find y_2 Or one can compute the value of dy/dx at the point

$$\left(x_0 + \frac{\Delta x}{2},\, y_0 + \frac{\Delta y}{2}\right)$$

and use this instead of the derivative at (x_0,y_0) in Eq. (7) to obtain an improved value of y_1, namely,

(9) $$y_1 = y_0 + \frac{dy}{dx}\bigg|_{\substack{x_0+\Delta x/2\\ y_0+\Delta y/2}}\Delta x = y_0 + f\left(x_0 + \frac{\Delta x}{2},\, y_0 + \frac{f(x_0,y_0)\,\Delta x}{2}\right)\Delta x$$

before continuing The procedure based on Eq (7) is known as **Euler's method**; that based on Eq (8) is known as the **modified Euler method**; and that based on Eq (9) is known as **Runge's method.**

In the Runge-Kutta method, three or more estimates of Δy are computed, and then the value of Δy which is finally used to determine y_1 is taken to be a linear combination of all of these Specifically, in **Kutta's third-order approximation** we let

$$\begin{aligned}
\Delta_1 y &\equiv k_1 = f(x_0,y_0)\,\Delta x \equiv f(x_0,y_0)h\\
\Delta_2 y &\equiv k_2 = f(x_0 + p\,\Delta x, y_0 + p\,\Delta_1 y)\,\Delta x \equiv f(x_0 + ph, y_0 + pk_1)h\\
\Delta_3 y &\equiv k_3 = f(x_0 + q\,\Delta x, y_0 + r\,\Delta_2 y + \overline{q-r}\,\Delta_1 y)\,\Delta x\\
&\qquad\qquad \equiv f(x_0 + qh, y_0 + rk_2 + \overline{q-r}k_1)h
\end{aligned}$$

and then put

$$\Delta y = ak_1 + bk_2 + ck_3$$

where a, b, c, p, q, r are constants to be determined to ensure the highest possible accuracy in Δy.

To do this, we must first expand k_1, k_2, k_3, and then Δy in terms of powers of $\Delta x \equiv h$, using implicit differentiation to compute the necessary derivatives Using subscripts to indicate the partial derivatives of f evaluated at $h = 0$, i e , letting

$$f_0 = f(x_0,y_0),\ f_1 = \frac{\partial f(x,y)}{\partial x}\bigg|_{x_0,y_0},\ f_2 = \frac{\partial f(x,y)}{\partial y}\bigg|_{x_0,y_0},\ f_{11} = \frac{\partial^2 f(x,y)}{\partial x^2}\bigg|_{x_0,y_0},\ \cdots$$

and using dk_i/dh and d^2k_i/dh^2 as abbreviations for $dk_i/dh\big|_{h=0}$ and $d^2k_i/dh^2\big|_{h=0}$, we thus have

$$
\begin{aligned}
k_1 &= f_0 h \\
k_2 &= \left\{ f_0 + \left(f_1 p + f_2 p \frac{dk_1}{dh} \right) h + \left[f_{11} p^2 + 2 f_{12} p^2 \frac{dk_1}{dh} + f_{22} p^2 \left(\frac{dk_1}{dh} \right)^2 \right] \frac{h^2}{2} + \qquad \right\} h \\
&= f_0 h + (f_1 + f_2 f_0) p h^2 + (f_{11} + 2 f_{12} f_0 + f_{22} f_0^2) \frac{p^2 h^3}{2} + \\
k_3 &= \left(f_0 + \left\{ f_1 q + f_2 \left[r \frac{dk_2}{dh} + (q - r) \frac{dk_1}{dh} \right] \right\} h \right. \\
&\quad + \left\{ f_{11} q^2 + 2 f_{12} q \left[r \frac{dk_2}{dh} + (q - r) \frac{dk_1}{dh} \right] + f_{22} \left[r \frac{dk_2}{dh} + (q - r) \frac{dk_1}{dh} \right]^2 \right. \\
&\qquad \left. \left. + f_2 \left[r \frac{d^2 k_2}{dh^2} + (q - r) \frac{d^2 k_1}{dh^2} \right] \right\} \frac{h^2}{2} + \qquad \right) h \\
&= f_0 h + (f_1 + f_2 f_0) q h^2 + [(f_{11} + 2 f_{12} f_0 + f_{22} f_0^2) q^2 + 2 f_2 (f_1 + f_2 f_0) p r] \frac{h^3}{2} + \cdots
\end{aligned}
$$

Hence, substituting,

$$
\begin{aligned}
\Delta y &= a k_1 + b k_2 + c k_3 \\
&= a f_0 h + b \left[f_0 h + (f_1 + f_2 f_0) p h^2 + (f_{11} + 2 f_{12} f_0 + f_{22} f_0^2) \frac{p^2 h^3}{2} + \qquad \right] \\
&\quad + c \left\{ f_0 h + (f_1 + f \quad) q h^2 + \left[(f_{11} + 2 f_{12} f_0 + f_{22} f_0^2) q^2 \right. \right. \\
&\qquad \left. \left. + 2 f_2 (f_1 + f_2 f_0) p r \right] \frac{h^3}{2} + \qquad \right\} \\
(10) \quad &= (a + b + c) f_0 h + (b p + c q)(f_1 + f_2 f_0) h^2 \\
&\quad + \left[\frac{b p^2 + c q^2}{2} (f_{11} + 2 f_{12} f_0 + f_{22} f_0^2) + c p r f_2 (f_1 + f_2 f_0) \right] h^3 +
\end{aligned}
$$

Now, since $y' = f(x,y)$, we have, by Maclaurin's expansion,

$$
\begin{aligned}
\Delta y &= y - y_0 = y_0' h + y_0'' \frac{h^2}{2!} + y_0''' \frac{h^3}{3!} + \\
(11) \quad &= f_0 h + (f_1 + f_2 f_0) \frac{h^2}{2!} + [(f_{11} + 2 f_{12} f_0 + f_{22} f_0^2) + f_2 (f_1 + f_2 f_0)] \frac{h^3}{3!} +
\end{aligned}
$$

Hence, Δy as given by (10) will agree with the Maclaurin expansion of Δy, as given by (11), through terms in $h^3 \equiv (\Delta x)^3$, provided

$$
(12) \qquad
\begin{aligned}
a + b + c &= 1 & \qquad b p + c q &= \frac{1}{2} \\
\frac{b p^2 + c q^2}{2} &= \frac{1}{6} & \qquad c p r &= \frac{1}{6}
\end{aligned}
$$

We thus have four equations in the six unknown parameters a, b, c, p, q, r The first three equations are linear in a, b, c and can easily be solved to express a, b, and c in terms of p and q Then the fourth equation can be used to express r in terms of p and q also

(13)
$$a = \frac{6pq - 3(p + q) + 2}{6pq} \qquad b = \frac{2 - 3q}{6p(p - q)}$$
$$c = \frac{2 - 3p}{6q(q - p)} \qquad r = \frac{q(q - p)}{p(2 - 3p)}$$

Since p and q are arbitrary, we thus have a two-parameter family of formulas which can be used for the step-by-step solution of the equation $y' = f(x,y)$ with an error which is of the order of $(\Delta x)^4 \equiv h^4$

The following particular cases are worthy of note.

(I)
$$a = \tfrac{1}{4} \qquad b = 0 \qquad c = \tfrac{3}{4} \qquad p = \tfrac{1}{3} \qquad q = r = \tfrac{2}{3}$$
$$\Delta y = \tfrac{1}{4}(k_1 + 3k_3)$$
where
$$k_1 = f(x_0,y_0)h$$
$$k_2 = f(x_0 + \tfrac{1}{3}h,\ y_0 + \tfrac{1}{3}k_1)h$$
$$k_3 = f(x_0 + \tfrac{2}{3}h,\ y_0 + \tfrac{2}{3}k_2)h$$

(II)
$$a = \tfrac{1}{4} \qquad b = c = \tfrac{3}{8} \qquad p = q = r = \tfrac{2}{3}$$
$$\Delta y = \tfrac{1}{8}(2k_1 + 3k_2 + 3k_3)$$
where
$$k_1 = f(x_0,y_0)h$$
$$k_2 = f(x_0 + \tfrac{2}{3}h,\ y_0 + \tfrac{2}{3}k_1)h$$
$$k_3 = f(x_0 + \tfrac{2}{3}h,\ y_0 + \tfrac{2}{3}k_2)h$$

The values of the parameters in case (II) cannot be obtained from Eqs (13), since $p = q$, but can be checked directly in Eqs (12)

The foregoing analysis can be extended without difficulty to yield step-by-step solution procedures in which the error is of the order of $(\Delta x)^5 \equiv h^5$ In particular, the following two sets of formulas are quite useful

(III)
$$\Delta y = \tfrac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$
where
$$k_1 = f(x_0,y_0)h$$
$$k_2 = f(x_0 + \tfrac{1}{2}h,\ y_0 + \tfrac{1}{2}k_1)h$$
$$k_3 = f(x_0 + \tfrac{1}{2}h,\ y_0 + \tfrac{1}{2}k_2)h$$
$$k_4 = f(x_0 + h,\ y_0 + k_3)h$$

(IV)
$$\Delta y = \tfrac{1}{8}(k_1 + 3k_2 + 3k_3 + k_4)$$
where
$$k_1 = f(x_0,y_0)h$$
$$k_2 = f(x_0 + \tfrac{1}{3}h,\ y_0 + \tfrac{1}{3}k_1)h$$
$$k_3 = f(x_0 + \tfrac{2}{3}h,\ y_0 + k_2 - \tfrac{1}{3}k_1)h$$
$$k_4 = f(x_0 + h,\ y_0 + k_3 - k_2 + k_1)h$$

The solution process based on (III) is often referred to specifically as *the* **Runge-Kutta method.**

Any of the Runge-Kutta formulas can be used to solve simultaneous and, hence, higher-order differential equations. For instance, using (III), we can tabulate the solution of the system of equations

$$\frac{dy}{dx} = f(x,y,z) \qquad \frac{dz}{dx} = g(x,y,z)$$

at intervals of $\Delta x \equiv h$ by computing

$$\begin{aligned}
\Delta_1 y &\equiv k_1 = f(x_0,y_0,z_0)h \\
\Delta_1 z &\equiv l_1 = g(x_0,y_0,z_0)h \\
\Delta_2 y &\equiv k_2 = f(x_0 + \tfrac{1}{2}h,\ y_0 + \tfrac{1}{2}k_1,\ z_0 + \tfrac{1}{2}l_1)h \\
\Delta_2 z &\equiv l_2 = g(x_0 + \tfrac{1}{2}h,\ y_0 + \tfrac{1}{2}k_1,\ z_0 + \tfrac{1}{2}l_1)h \\
\Delta_3 y &\equiv k_3 = f(x_0 + \tfrac{1}{2}h,\ y_0 + \tfrac{1}{2}k_2,\ z_0 + \tfrac{1}{2}l_2)h \\
\Delta_3 z &\equiv l_3 = g(x_0 + \tfrac{1}{2}h,\ y_0 + \tfrac{1}{2}k_2,\ z_0 + \tfrac{1}{2}l_2)h \\
\Delta_4 y &\equiv k_4 = f(x_0 + h,\ y_0 + k_3,\ z_0 + l_3)h \\
\Delta_4 z &\equiv l_4 = g(x_0 + h,\ y_0 + k_3,\ z_0 + l_3)h
\end{aligned}$$

and then using the formulas

$$\begin{aligned}
\Delta y &= \tfrac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\
\Delta z &= \tfrac{1}{6}(l_1 + 2l_2 + 2l_3 + l_4)
\end{aligned}$$

If the various increments are computed in the indicated order, each involves only quantities which have previously been calculated.

EXAMPLE 1

Tabulate the solution of $y' = x^2 + y$ at intervals of $h = 0.1$ if $y = -1$ when $x = 0$.

Using the Runge-Kutta formulas (III) for the first increment, we have

$$k_1 = -0.1000 \qquad k_2 = -0.1048 \qquad k_3 = -0.1050 \qquad k_4 = -0.1095$$

and $\quad \Delta y = -0.1048$

Hence, $\quad y_1 = y_0 + \Delta y = -1.1048$

For the second increment, we have, similarly,

$$k_1 = -0.1095 \qquad k_2 = -0.1137 \qquad k_3 = -0.1139 \qquad k_4 = -0.1179$$

and $\quad \Delta y = -0.1138$

Hence, $\quad y_2 = y_1 + \Delta y = -1.2186$

For the third increment, we have

$$k_1 = -0.1179 \qquad k_2 = -0.1215 \qquad k_3 = -0.1217 \qquad k_4 = -0.1250$$

and $\quad \Delta y = -0.1216$

Hence, $\quad y_3 = y_2 + \Delta y = -1.3402$

This process can, of course, be continued as far as desired. However, we shall calculate just one more value of y, this time using Milne's method, which can now be applied since we have values for y_0, y_1, y_2, and y_3. To do this we must first compute y_0', y_1', y_2', y_3' from the differential equation, getting

$$y'_0 = -1.000 \qquad y'_1 = -1.0948 \qquad y'_2 = -1.1786 \qquad y'_3 = -1.2502$$

With these values, we can now use the open formula, Eq. (3), to obtain

$$y_4 = -1.4682$$

Using this value, we find from the differential equation that

$$y'_4 = -1.3082$$

With this we can use Eq. (3) again to find y_5, or we can first use the closed formula, Eq. (6), to check y_4 before continuing. In this case Eq. (6) also gives us $y_4 = -1.4682$, which is a good check on the accuracy of our calculations.

The differential equation $y' = x^2 + y$ is so simple that it can be solved exactly without recourse to numerical methods, and by the methods of Chap. 1 (or Chap. 2) we find at once that the required solution is

$$y = e^x - x^2 - 2x - 2$$

For $x = 1, 2, 3, 4$ this gives us the correct values

$$y_1 = -1.1048 \qquad y_2 = -1.2186 \qquad y_3 = 1.3401 \qquad y_4 = -1.4682$$

The values we computed for y_1, y_2, and y_4 agree with these to four decimal places, and the value we computed for y_3 differs from the correct value by only 1 in the fourth place.

EXERCISES

1 Using the Runge-Kutta method (III), find y_5 and y_9 in Example 1 without finding y_6, y_7 or y_8. How do these values compare with the correct values?

2 Using Kutta's third-order approximation (I), tabulate the solution of the equation $y' = x - y$ at intervals of $h = 0.1$ if $y = 1$ when $x = 1$.

3 Using Kutta's third-order approximation (II), tabulate the solution of the equation $y' = x + y$ at intervals of $h = 0.1$ if $y = 1$ when $x = 0$.

4 Using Milne's method, tabulate the solution of the equation $y' = x + y$ at intervals of $h = 0.1$ if $y = 1$ when $x = 0$.

5 Using the Runge-Kutta method (III), find y_1 and z_1 for the solution of the system

$$\frac{dy}{dx} = x + z \qquad \frac{dz}{dx} = x - y$$

given $h = 0.1$, and $y = 0$, $z = 1$ when $x = 0$.

6 Using Milne's method, tabulate the solution of the system

$$\frac{dy}{dx} = y^2 + xz \qquad \frac{dz}{dx} = x^2 + yz$$

at intervals of $h = 0.1$, given $y_0 = 0$, $z_0 = 1$.

7 Work Exercise 2 using the open formula

$$y_{n+1} = y_n + h(y'_n + \tfrac{1}{2}\,\Delta y'_{n-1} + \tfrac{5}{12}\,\Delta^2 y'_{n-2} + \tfrac{3}{8}\,\Delta^3 y'_{n-3} + \tfrac{251}{720}\,\Delta^4 y'_{n-4})$$

and the closed formula

$$y_{n+1} = y_n + h(y'_{n+1} - \tfrac{1}{2}\,\Delta y'_n - \tfrac{1}{12}\,\Delta^2 y'_{n-1} - \tfrac{1}{24}\,\Delta^3 y'_{n-2} - \tfrac{19}{720}\,\Delta^4 y'_{n-3})$$

(These equations constitute the so-called **Adams-Bashforth method** for the numerical solution of differential equations.)

8 Explain how the Adams-Bashforth method described in Exercise 7 can be extended to systems of differential equations and equations of higher order.

9 Eliminate the differences from the formulas of Exercise 7, and express y_{n+1} directly in terms of y_n and the various values of y'.

10 Work Exercise 3 using the open formula

$$y_{n+1} = y_n + h(\tfrac{23}{12}y'_n - \tfrac{16}{12}y_{n-1} + \tfrac{5}{12}y'_{n-2})$$

and the closed formula

$$y_{n+1} = y_n + h(\tfrac{5}{12}y'_{n+1} + \tfrac{8}{12}y'_n - \tfrac{1}{12}y'_{n-1})$$

(These equations constitute the so-called **Adams-Moulton method** for the numerical solution of differential equations)

11 Using the open formula

$$y_{n+1} = 2y_n - y_{n-1} + h^2(y''_n + \tfrac{1}{12}\,\Delta^2 y''_{n-2})$$

and the closed formula

$$y_{n+1} = 2y_n - y_{n-1} + h^2(y''_n + \tfrac{1}{12}\,\Delta^2 y''_{n-1})$$

tabulate the solution of the equation $y'' = x + y$ at intervals of $\Delta x = h = 0\,1$, given $y_0 = 1$, $y'_0 = 0$ How do your results compare with the exact solution?

12 Set up the Kutta third-order approximation corresponding to the values $p = \tfrac{1}{2}$, $q = 1$, and show that it reduces to Simpson's rule when $f(x,y)$ is independent of y

13 By expanding each term in Eq (3) around the point $x = x_{n-1}$, show that the principal part of the error in Milne's open formula is $\tfrac{14}{45}h^5y^{v}_{n-1}$ What is the principal part of the error in Milne's closed formula?

14 Find the equation of the polynomial of minimum degree for which y and y' take on prescribed values (y_0,y'_0) and (y_1,y'_1) at $x = 0$ and $x = h$ What is the value of y_2 given by this polynomial? How might this result be used to carry out the step-by-step integration of a differential equation of the form $y' = f(x,y)$? How might an accompanying closed formula be obtained?

15 Find the equation of the polynomial of minimum degree for which y and y'' take on prescribed values (y_0,y''_0) and (y_1,y''_1) at $x = 0$ and $x = h$ What is the value of y_2 given by this polynomial? How might this result be used to carry out the step-by-step integration of a differential equation of the form $y'' = f(x,y)$? How might an accompanying closed formula be obtained?

4.5

Difference equations

The many similarities we have already observed between the calculus of finite differences and the ordinary, or infinitesimal, calculus suggest that there should be a theory of *difference equations* roughly paralleling the theory of *differential equations*, and this is indeed the case However, in the study of difference equations we do not ordinarily consider equations of the form

$$f(\Delta)y = \phi(x) \tag{1}$$

as might be expected by analogy with the differential equation

$$f(D)y = \phi(x) \tag{2}$$

but rather equations of the form

$$f(E)y = \phi(x) \tag{3}$$

This, of course, is simply a matter of notational convenience, since, using the operational equivalence $\Delta = E - 1$, any function of Δ can be transformed at once into a function of E, and vice versa In this section we shall restrict ourselves to the case of a single **linear, constant-coefficient difference equation**

$$(a_0E^r + a_1E^{r-1} + \cdots + a_{r-1}E + a_r)y = \phi(x) \tag{4}$$

where $\phi(x)$ is a linear combination of terms or products of terms from the set

k^x $\cos kx$ $\sin kx$ k a constant

and x^n n a nonnegative integer

Since the substitution $t = hx$ will transform a function of t tabulated at intervals of h into a function of x tabulated at unit intervals, it is clearly no restriction to assume $h = 1$, so that invariably $Ef(x) = f(x + 1)$, and we shall do this throughout the present section We shall base our solution of Eq (4) primarily on analogy with linear, constant-coefficient differential equations, and such theoretical results as we may need we shall merely quote without proof

In Eq (4), if both a_0 and a_r are different from zero, as we shall henceforth suppose, the positive integer r is called the **order** of the equation. If $\phi(x)$ is identically zero, Eq (4) is said to be **homogeneous**; if $\phi(x)$ is not identically zero, Eq (4) is said to be **nonhomogeneous.** By a **solution** of (4) we mean a function of x with the property that, when it is substituted into (4), it reduces the equation to an identity From a theoretical point of view both x and y should be regarded as continuous variables related by Eq (4) on a set of equally spaced values of x However, in practical problems we are almost always interested in y only for the discrete values $x = \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$, and in our work we shall attempt no more than the determination of solutions defined on this range

For the second-order linear difference equation, with either variable or constant coefficients, we have three theorems completely analogous to the fundamental theorems of Sec 2 1

THEOREM 1

If $y_1(x)$ and $y_2(x)$ are any two solutions of the homogeneous equation

$$(a_0E^2 + a_1E + a_2)y = 0$$

then $c_1y_1(x) + c_2y_2(x)$, where c_1 and c_2 are arbitrary constants, is also a solution

THEOREM 2

If $y_1(x)$ and $y_2(x)$ are two solutions of the homogeneous equation

$$(a_0E^2 + a_1E + a_2)y = 0$$

for which

$$C[y_1(x), y_2(x)]\dagger = \begin{vmatrix} y_1(x) & y_2(x) \\ Ey_1(x) & Ey_2(x) \end{vmatrix} \neq 0$$

then any solution $y_3(x)$ of the homogeneous equation can be written in the form $y_3(x) = c_1y_1(x) + c_2y_2(x)$, where c_1 and c_2 are suitable constants.

As a consequence of Theorem 2, the expression $c_1y_1(x) + c_2y_2(x)$ is called a **complete solution** of the homogeneous equation when the particular solutions $y_1(x)$ and $y_2(x)$ satisfy the condition

$$C[y_1(x), y_2(x)] \neq 0$$

THEOREM 3

If $Y(x)$ is any solution of the nonhomogeneous equation

$$(a_0E^2 + a_1E + a_2)y = \phi(x)$$

and if $c_1y_1(x) + c_2y_2(x)$ is a complete solution of the homogeneous equation obtained from this by deleting the term $\phi(x)$, then

$$y = c_1y_1(x) + c_2y_2(x) + Y(x)$$

is a complete solution of the nonhomogeneous equation

As in the theory of differential equations, any complete solution of the related homogeneous equation is usually called a **complementary function** of the nonhomogeneous equation The extension of these theorems to difference equations of order greater than 2 is obvious

To find particular solutions of the homogeneous equation

$$(a_0E^2 + a_1E + a_2)y = 0 \tag{5}$$

when the coefficients a_0, a_1, a_2 are constants, we might try, as with the analogous differential equation,

$$y = e^{mx} \tag{6}$$

However, it is more convenient to assume

$$y = M^x \tag{7}$$

which is clearly equivalent to (6) with $M = e^m$ Substituting this into (5), recalling our agreement that $Ef(x) = f(x + 1)$, we obtain

$$a_0M^{x+2} + a_1M^{x+1} + a_2M^x = 0$$

or, dividing out M^x,

$$a_0M^2 + a_1M + a_2 = 0 \tag{8}$$

---

† The function $C[y_1(x), y_2(x)]$ is customarily referred to as **Casorati's determinant**, after the Italian mathematician Felica Casorati (1835–1890) Its resemblance to the Wronskian $W[y_1(x), y_2(x)]$ (see Sec 2.1) is apparent

Naturally enough, this is called the **characteristic equation** of the difference equation (5).

If the roots M_1 and M_2 of (8) are distinct, then

$$C(M_1^x,M_2^x) = M_1^xM_2^{x+1} - M_2^xM_1^{x+1} = M_1^xM_2^x(M_2 - M_1) \neq 0\dagger$$

and, hence, by Theorem 2, a complete solution of Eq (5) is

$$y = c_1M_1^x + c_2M_2^x \tag{9}$$

If M_1 and M_2 are real, this is a completely acceptable form of the solution However, if M_1 and M_2 are complex, then (9) is inconvenient for most purposes, and it is desirable that we reduce it to a more useful form To do this, let the roots be

$$M_1,\ M_2 = p \pm iq = re^{\pm i\theta}$$

where $\quad r = \sqrt{p^2 + q^2} \quad$ and $\quad \tan\theta = \frac{q}{p}$‡

Then we can write

$$\begin{aligned} y &= c_1(re^{i\theta})^x + c_2(re^{-i\theta})^x \\ &= r^x(c_1e^{i\theta x} + c_2e^{-i\theta x}) \\ &= r^x[c_1(\cos\theta x + i\sin\theta x) + c_2(\cos\theta x - i\sin\theta x)] \\ &= r^x[(c_1 + c_2)\cos\theta x + i(c_1 - c_2)\sin\theta x] \end{aligned}$$

or, renaming the constants,

$$y = r^x(A\cos\theta x + B\sin\theta x) \tag{10}$$

If $M_1 = M_2$, clearly $C(M_1^x,M_2^x) = 0$, and we must find a second, independent solution before we can construct the complete solution of (5) Again by analogy with differential equations, we are led to try

$$y = xM_1^x$$

and we find by direct substitution that this is indeed a solution when the characteristic equation (8) has equal roots For we have

$$\begin{aligned} &a_0(x + 2)M_1^{x+2} + a_1(x + 1)M_1^{x+1} + a_2xM_1^x \\ &\qquad = xM_1^x(a_0M_1^2 + a_1M_1 + a_2) + M_1^{x+1}(2a_0M_1 + a_1) = 0 \end{aligned}$$

since the coefficient of xM_1^x vanishes, because in any case M_1 satisfies the characteristic equation (8), and the coefficient of M_1^{x+1} vanishes, because when the characteristic equation has equal roots their common value is $M_1 = -a_1/2a_0$ Moreover, for the solutions M_1^x and xM_1^x we have

$$C(M_1^x,xM_1^x) = M_1^x(x + 1)M_1^{x+1} - xM_1^xM_1^{x+1} = M_1^{2x+1} \neq 0$$

Hence, according to Theorem 2, a complete solution when the

---

† Since $a_2 \neq 0$, or else the difference equation would be of order less than 2, contrary to hypothesis, it is clear that neither M_1 nor M_2 can be zero.

‡ For a discussion of the exponential form of a complex number, see Sec 14 7

characteristic equation has equal roots is

(11) $$y = c_1M_1^x + c_2xM_1^x$$

The results of the preceding discussion are summarized in Table 4.2

table 4 2

Difference equation $(a_0E^2 + a_1E + a_2)y = 0 \qquad a_0, a_2 \neq 0$
Characteristic equation $a_0M^2 + a_1M + a_2 = 0$

Nature of the roots of the characteristic equation	Condition on the coefficients of the characteristic equation	Complete solution of the difference equation
Real and unequal $M_1 \neq M_2$	$a_1^2 - 4a_0a_2 > 0$	$y = c_1M_1^x + c_2M_2^x$
Real and equal $M_1 = M_2$	$a_1^2 - 4a_0a_2 = 0$	$y = c_1M_1^x + c_2xM_1^x$
Conjugate complex $M_1 = p + iq$ $M_2 = p \quad iq$	$a_1^2 - 4a_0a_2 < 0$	$y = r^x(A\cos\theta x + B\sin\theta x)$ $r = \sqrt{p^2 + q^2}$ $\tan\theta = q/p$

EXAMPLE 1

Find a complete solution of the difference equation $(E^2 + 2E + 4)y = 0$

The characteristic equation in this case is $M^2 + 2M + 4 = 0$, and its roots are M_1, $M_2 = -1 \pm i\sqrt{3}$ Since

$$r = \sqrt{(-1)^2 + (\sqrt{3})^2} = 2 \qquad \text{and} \qquad \theta = \tan^{-1}\frac{\sqrt{3}}{-1} = \frac{2\pi}{3}$$

we have as a complete solution

$$y = 2^x\left(A\cos\frac{2\pi x}{3} + B\sin\frac{2\pi x}{3}\right)$$

To solve the nonhomogeneous equation

(12) $$(a_0E^2 + a_1E + a_2)y = \phi(r)$$

we must, according to Theorem 3, add a particular solution of (12) to a complete solution of the related homogeneous equation (5). To find the necessary particular solution Y, we use the method of undetermined coefficients, starting with an arbitrary linear combination of all the independent terms which arise from $\phi(x)$ by repeatedly applying the operator E As in the case of differential equations, if any term in the initial choice for Y duplicates a term in the complementary function, it and all associated terms must be multiplied by x until duplication is eliminated. The procedure is summarized in Table 4 3

table 4.3

Difference equation $(a_0E^2 + a_1E + a_2)y = \phi(x)$

$\phi(x)$*	Necessary choice for particular solution Y†
1 a (constant)	A
2 ax^k (k a positive integer)	$A_0x^k + A_1x^{k-1} + \cdots + A_{k-1}x + A_k$
3. ak^x	Ak^x
4 $a \cos kx$	$A \cos kx + B \sin kx$
5 $a \sin kx$	
6 $ax^kl^x \cos mx$	$(A_0x^k + \cdots + A_{k-1}x + A_k)l^x \cos mx$ $+ (B_0x^k + \cdots + B_{k-1}x + B_k)l^x \sin mx$
7 $ax^kl^x \sin mx$	

* When $\phi(x)$ consists of a sum of several terms, the appropriate choice for Y is the sum of the Y expressions corresponding to these terms individually

† Whenever a term in any of the Y's listed in this column duplicates a term already in the complementary function, all terms in that Y must be multiplied by the lowest positive integral power of x sufficient to eliminate the duplication

EXAMPLE 2

Find a complete solution of the difference equation $(E^2 - 5E + 6)y = x + 2^x$

The characteristic equation in this case is $M^2 - 5M + 6 = 0$, and from its roots $M_1 = 2$, $M_2 = 3$, we can immediately construct the complementary function $y = c_12^x + c_23^x$ For a particular solution we would ordinarily try $Y = Ax + B + C2^x$ However, it is clear that $C2^x$ duplicates a term in the complementary function Hence, we must multiply $C2^x$ by x before incorporating it in our choice for Y Thus we substitute $Y = Ax + B + Cx2^x$ into the difference equation, getting

$$[A(x + 2) + B + C(x + 2)2^{x+2}] - 5[A(x + 1) + B + C(x + 1)2^{x+1}]$$
$$+ 6[Ax + B + Cx2^x] = x + 2^x$$

or

$$2Ax + (-3A + 2B) - 2C2^x = x + 2^x$$

which will be an identity if and only if $A = \frac{1}{2}$, $B = \frac{3}{4}$, and $C = -\frac{1}{2}$ A complete solution is, therefore,

$$y = c_12^x + c_23^x + \frac{2x + 3}{4} - \frac{x2^x}{2}$$

EXAMPLE 3

Find the sum of the series

$$s = \sum_{x=1}^{n} xk^x \qquad k \neq 1$$

Clearly s satisfies the first-order difference equation

$$s_{n+1} - s_n \equiv (E - 1)s_n = (n + 1)k^{n+1}$$

The characteristic equation here is $M - 1 = 0$, and so the complementary function is simply $s = c_1(1)^n = c_1$. To find a particular integral, we assume

$$S = (an + b)k^{n+1}$$

Then, substituting, we must have

$$[a(n + 1) + b]k^{n+2} - (an + b)k^{n+1} = (n + 1)k^{n+1}$$

or, dividing out k^{n+1} and collecting terms,

$$n(ak - a) + (ak + bk - b) = n + 1$$

This will be an identity in the variable n if and only if

$$a(k - 1) = 1 \qquad \text{and} \qquad ak + b(k - 1) = 1$$

or

$$a = \frac{1}{k - 1} \qquad \text{and} \qquad b = -\frac{1}{(k - 1)^2}$$

Hence,

$$S = \left[\frac{n}{k - 1} - \frac{1}{(k - 1)^2}\right]k^{n+1}$$

and a complete solution is

$$s = c_1 + S = c_1 + \frac{n(k - 1) - 1}{(k - 1)^2}k^{n+1}$$

To determine c_1 we use the obvious fact that when $n = 1$, $s = k$. Thus we must have

$$k = c_1 + \frac{k - 2}{(k - 1)^2}k^2 \qquad \text{or} \qquad c_1 = \frac{k}{(k - 1)^2}$$

Hence, finally,

$$s = \frac{k + [n(k - 1) - 1]k^{n+1}}{(k - 1)^2}$$

EXAMPLE 4

In the system shown in Fig. 4.2a the point P_0 is kept at the constant potential V_0 with respect to the ground. What is the potential at each of the points $P_1, P_2, \ldots, P_{n-1}$?

According to Kirchhoff's first law, the sum of the currents flowing toward any junction in a network must equal the sum of the currents flowing away from that junction. Hence, at a general point P_{x+1} (Fig. 4.2b) we have

$$i_x = i_{x+1} + I_{x+1}$$

or, replacing each current by its equivalent according to Ohm's law,

$$\frac{V_x - V_{x+1}}{r} = \frac{V_{x+1} - V_{x+2}}{r} + \frac{V_{x+1}}{2r}$$

or, finally,

$$V_{x+2} - \tfrac{5}{2}V_{x+1} + V_x = 0 \tag{13}$$

This equation holds for $x = 2, \ldots, n - 2$, that is, at all but the points P_1 and P_{n-1}, where we have the respective conditions

$$-V_2 + \tfrac{5}{2}V_1 = V_0 \qquad \text{since } V_0 \text{ is given} \tag{14}$$

$$-\tfrac{5}{2}V_{n-1} + V_{n-2} = 0 \qquad \text{since } V_n = 0 \tag{15}$$

Equations (13), (14), and (15) constitute a system of $n - 1$ linear equations from which the unknown potentials $V_1, V_2, \ldots, V_{n-1}$ can be found by completely elementary though very tedious steps for any particular value of n. However, it is much simpler and more elegant to regard Eq. (13) as a second-order difference equation, subject to the end conditions (14) and (15), which will serve to determine the values of the arbitrary constants appearing in any complete solution of (13).

Taking this point of view, we first set up the characteristic equation of Eq. (13)

$$M^2 - \tfrac{5}{2}M + 1 = 0$$

From its roots $M_1 = \tfrac{1}{2}$ and $M_2 = 2$, we then construct a complete solution of Eq. (13), namely,

$$V_x = A(\tfrac{1}{2})^x + B2^x$$

Substituting this into Eqs. (14) and (15), we have

$$-\left(\frac{A}{4} + 4B\right) + \frac{5}{2}\left(\frac{A}{2} + 2B\right) = V_0$$

$$-\frac{5}{2}\left(\frac{A}{2^{n-1}} + B2^{n-1}\right) + \frac{A}{2^{n-2}} + B2^{n-2} = 0$$

or

$$A + B = V_0 \qquad \text{and} \qquad \frac{A}{2^n} + B2^n = 0$$

from which we find at once

$$A = \frac{2^{2n}}{2^{2n} - 1} V_0 \qquad B = -\frac{1}{2^{2n} - 1} V_0$$

The final solution is, therefore,

$$V_x = \left(\frac{2^{2n}}{2^x} - 2^x\right)\frac{V_0}{2^{2n} - 1}$$

That this reduces to V_0 when $x = 0$ and reduces to 0 when $x = n$ is easily verified.

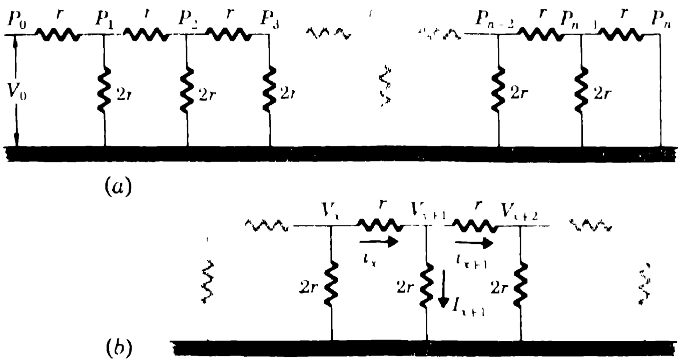


FIGURE 4.2

A ladder-type network with identical loops. (Although the network shown in Fig. 4.2a appears to contain exactly seven loops, the number of loops is actually indefinite. This is implied by the fact that the central portion of the figure is drawn with lighter lines; this convention will be used throughout the book to suggest a configuration of indefinite extent.)

EXERCISES

1 Find a complete solution of each of the following equations

a $(E^2 + 7E + 12)y = 0$ **b** $(E^2 + 6E + 9)y = 0$

c $(E^2 + 2E + 2)y = 0$ **d** $(\Delta^2 - 3\Delta + 2)y = 0$

2 Find a complete solution of each of the following equations

a $(E^2 - E - 6)y = x^2$ **b** $(4E^2 - 4E + 1)y = x + 2 + 2^x$

c $(E^2 + 4)y = \cos x$ **d** $(\Delta^2 + 6\Delta + 18)y = 2^{-x}$

e $(E^2 - 3E + 2)y = 2^x + 2^{-x}$ **f** $(E^2 - 4E + 4)y = 2^x$

3 Find a complete solution of each of the following equations

a $(E^2 - E - 6)y = x + 3^x$ **b** $(E^2 + 1)y = \sin x$

4 Find a complete solution of each of the following equations

a $(E^3 - 6E^2 + 11E - 6)y = 0$ **b** $(E^4 - 16)y = x + 3^x$

c $(E^4 + 10E^2 + 9)y = 0$ **d** $(E^4 + 8E^2 - 9)y = 5$

5 Show that the difference equation $(E^2 - 2\lambda E + 1)y = 0$ has the indicated solution in each of the following special cases

$$
\begin{array}{lll}
\lambda < -1 & y = A(-1)^x \cosh \mu x + B(-1)^x \sinh \mu x & \cosh \mu = -\lambda \\
\lambda = -1 & y = A(-1)^x + Bx(-1)^x & \\
-1 < \lambda < 1 & y = A \cos \mu x + B \sin \mu x & \cos \mu = \lambda \\
\lambda = 1 & y = A + Bx & \\
1 < \lambda & y = A \cosh \mu x + B \sinh \mu x & \cosh \mu = \lambda
\end{array}
$$

6 Work Example 4 with both P_0 and P_n maintained at the constant potential V_0

7 Work Example 4 given that the common value of the resistances in the vertical branches is kr

8 A system consists of n spring-connected masses, as shown in Fig 4 3 What is the displacement of each mass from its original position when the system is again in equilibrium after a force F_0 is applied to the right-hand end?

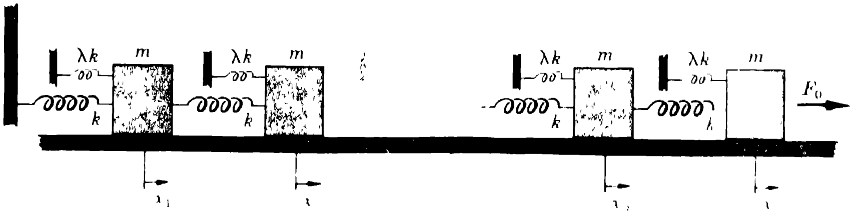


FIGURE 4.3

(See explanation of convention used for Fig 4 2)

9 Show that the nth-order determinant

$$
D_n = \begin{vmatrix}
\lambda & 1 & 0 & & 0 & 0 \\
1 & \lambda & 1 & & 0 & 0 \\
0 & 1 & \lambda & & 0 & 0 \\
 & & & & & \\
0 & 0 & 0 & \cdot & \lambda & 1 \\
0 & 0 & 0 & & 1 & \lambda
\end{vmatrix}
$$

satisfies the difference equation $(E^2 - \lambda E + 1)D = 0$ Hence show that, when $\lambda > 2$,

$$D_n = \frac{\sinh (n+1)\mu}{\sinh \mu} \qquad \text{where } \cosh \mu = \frac{\lambda}{2}$$

What is D_n if $\lambda = 2$? $-2 < \lambda < 2$? $\lambda = -2$? $\lambda < -2$?

10 If $y_1(x)$ and $y_2(x)$ are any two solutions of the general linear second-order difference equation $[a_0(x)E^2 + a_1(x)E + a_2(x)]y = 0$, show that Casorati's determinant $C[y_1(x),y_2(x)]$ satisfies the relation $[a_0(x)E - a_2(x)]C = 0$ [Hint Write down the conditions that both $y_1(x)$ and $y_2(x)$ satisfy the given equation, then eliminate the terms in $Ey_1(x)$ and $Ey_2(x)$]

11 Prove Theorem 1

12 Prove Theorem 2 in the special case where the coefficients are constants (Hint Recall the proof of Theorem 2, Sec 2 1, and use the result of Exercise 10)

13 Prove Theorem 3

14 Show that the integral $I_n = \int_0^\pi \frac{\cos nt - \cos n\lambda}{\cos t - \cos \lambda}\,dt$ satisfies the equation

$$[E^2 - (2\cos\lambda)E + 1]I = 0$$

Solve this equation, and find an explicit expression for I_n

15 Discuss the solution of each of the following equations

a $Ey = 0$

b $Ey = \phi(x)$

c $(a_0E^2 + a_1E)y = 0$

d $(a_0E^2 + a_1E)y = \phi(x)$

4.6

The method of least squares

The problem of curve fitting admits of two somewhat different interpretations In the first place, we may ask for the equation of a curve of prescribed type which passes exactly through each point of a given set For polynomial curves this is most easily accomplished by means of interpolation formulas such as we developed in Sec 4 2 On the other hand we may weaken these requirements and ask for some simpler curve whose equation contains too few parameters to permit it to pass exactly through each given point but which comes "as close as possible" to each point. For instance, given a set of points as in Fig 4 4a, a straight line passing as close as possible to each point may very well be more useful than some complicated curve passing exactly through

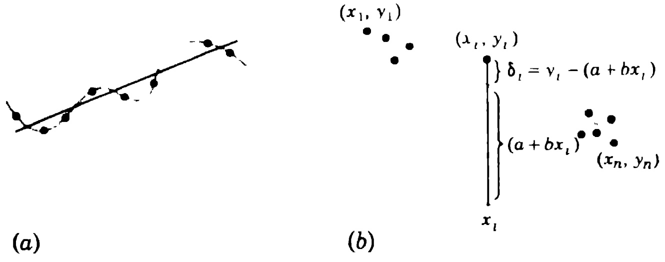


FIGURE 4 4
The approximate fitting of a straight line to a set of points

each point This will certainly be the case with experimental data which theoretically should fall along a straight line but which fail to do so because of errors of observation The necessary measure of "as close as possible" is almost universally taken to be the least-square criterion,* and the process of applying this criterion is known as the **method of least squares,** which we shall now develop

Let us begin by supposing that we wish to fit a straight line l whose equation is

(1) $$y = a + bx$$

to the n points (x_1,y_1), (x_2,y_2), , (x_n,y_n) Since two points completely determine a straight line, it will in general be impossible for the required line to pass through more than two of the given points, and it may not pass through any. Hence, the coordinates of the general point (x_i,y_i) will not satisfy Eq (1). That is, when we substitute x_i into Eq (1), we get, not y_i, but rather the ordinate of l, which, as we see in Fig 4 4b, differs from y_i by δ_i In other words,

(2) $$y_i - (a + bx_i) = \delta_i \neq 0$$

If we compute the discrepancy δ_i for each point of the set and form the sum of the squares of these quantities (in order to prevent large positive and large negative δ's canceling each other and thereby giving an unwarranted impression of accuracy), we obtain

(3) $$E = \sum_{i=1}^{n} \delta_i^2 = (y_1 - a - bx_1)^2 + (y_2 - a - bx_2)^2 + \\ + (y_n - a - bx_n)^2$$

The quantity E is obviously a measure of how well the line l fits the set of points as a whole For E will be zero if and only if each of the points lies on l, and the larger E is, the farther the points are, on the average, from l The least-square criterion is now simply this that *the parameters a and b should be chosen so as to make the sum of the squares of the deviations E as small as possible*

To do this, we apply the usual conditions for minimizing a function of several variables and equate to zero the two first partial derivatives, $\frac{\partial E}{\partial a}$ and $\frac{\partial E}{\partial b}$ This gives us the two equations

$$\frac{\partial E}{\partial a} = 2(y_1 - a - bx_1)(-1) + 2(y_2 - a - bx_2)(-1) + \\ + 2(y_n - a - bx_n)(-1) = 0$$

$$\frac{\partial E}{\partial b} = 2(y_1 - a - bx_1)(-x_1) + 2(y_2 - a - bx_2)(-x_2) + \\ + 2(y_n - a - bx_n)(-x_n) = 0$$

---

* A brief discussion of the reasons for this will be found in A M Mood, "Introduction to the Theory of Statistics," p 311, McGraw-Hill Book Company, New York, 1950.

or, dividing by 2 and collecting terms on the unknown coefficients a and b,

$$na + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{4}$$

$$a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \tag{5}$$

Equations (4) and (5) are two simultaneous linear equations whose solution for a and b presents no difficulty

For $i = 1, 2, \ldots, n$, (2) defines a system of n equations in the two unknowns a and b which should, ideally, be satisfied, but which actually are not Moreover, minimizing E is nothing more than minimizing the sum of the squares of the amounts by which these n equations fail to be satisfied

The preceding observation suggests a somewhat more general point of view, namely, that the method of least squares is simply a process for finding the best possible values for a set of m unknowns, say $x_1, x_2, \ldots, x_m$, connected by n linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= b_2 \\ &\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= b_n \end{aligned}$$

when $n > m$ Since the number of equations exceeds the number of unknowns, the system presumably does not admit of an exact solution, i e , there is no set of values for $x_1, x_2, \ldots, x_m$ for which each equation is exactly satisfied Hence, we consider the discrepancies

$$\delta_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{im}x_m - b_i \neq 0 \qquad i = 1, 2, \ldots, n$$

and attempt to find values for $x_1, x_2, \ldots, x_m$ which will make

$$E = \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} (a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{im}x_m - b_i)^2$$

as small as possible

To minimize E we must equate to zero each of its first partial derivatives

$$\frac{\partial E}{\partial x_1}, \quad \frac{\partial E}{\partial x_2}, \quad \ldots, \quad \frac{\partial E}{\partial x_m}$$

For $\dfrac{\partial E}{\partial x_1}$ this gives the equation

$$\frac{\partial E}{\partial x_1} = \sum_{i=1}^{n} 2(a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{im}x_m - b_i)(a_{i1}) = 0$$

or

$$x_1 \sum_{i=1}^{n} a_{i1}a_{i1} + x_2 \sum_{i=1}^{n} a_{i1}a_{i2} + \cdots + x_m \sum_{i=1}^{n} a_{i1}a_{im} = \sum_{i=1}^{n} a_{i1}b_i$$

and similarly, for the other partial derivatives,

$$x_1 \sum_{i=1}^{n} a_{i2}a_{i1} + x_2 \sum_{i=1}^{n} a_{i2}a_{i2} + \cdots + x_m \sum_{i=1}^{n} a_{i2}a_{im} = \sum_{i=1}^{n} a_{i2}b_i$$

$$\cdots \qquad \cdots \qquad \cdots \qquad \cdots$$

$$x_1 \sum_{i=1}^{n} a_{im}a_{i1} + x_2 \sum_{i=1}^{n} a_{im}a_{i2} + \cdots + x_m \sum_{i=1}^{n} a_{im}a_{im} = \sum_{i=1}^{n} a_{im}b_i$$

We have thus obtained a system of m linear equations in the m unknowns $x_1, x_2, \ldots, x_m$, whose solution is now a routine matter. As a practical detail, it is worthy of note that these minimizing conditions, or **normal equations,** as they are usually called, can be written down at once according to the following rule:

rule 1 If each of n linear equations in the m unknowns

$$x_1, x_2, \ldots, x_m \qquad n > m$$

is multiplied by the coefficient of x_i in that equation, the sum of the resulting equations is the ith normal equation in the least-square solution of the system.

EXAMPLE 1

By the method of least squares, fit a parabolic equation $y = a + bx + cx^2$ to the data

x	-3	-2	0	3	4
y	18	10	2	2	5

Substituting these pairs of values into the equation $y = a + bx + cx^2$, we find that a, b, and c should satisfy the conditions

$$\begin{aligned} a - 3b + 9c &= 18 \\ a - 2b + 4c &= 10 \\ a &= 2 \\ a + 3b + 9c &= 2 \\ a + 4b + 16c &= 5 \end{aligned}$$

In general, three unknowns cannot be made to satisfy more than three conditions; hence, the most we can do is to determine values of a, b, and c which will satisfy these equations as nearly as possible.

To set up the first of the three normal equations required by the method of least squares, we must multiply each of the equations of condition by the coefficient of a in that equation and add, getting in this case simply the sum of the five equations

$$5a + 2b + 38c = 37$$

To set up the second normal equation, we multiply each equation by the coefficient of b in that equation and add, getting

$$\begin{aligned} -3a + 9b - 27c &= -54 \\ -2a + 4b - 8c &= -20 \\ 0 + 0 + 0 &= 0 \\ 3a + 9b + 27c &= 6 \\ 4a + 16b + 64c &= 20 \\ \hline 2a + 38b + 56c &= -48 \end{aligned}$$

In the same way, multiplying each equation by the coefficient of c in that equation, we get the third normal equation

$$\begin{aligned} 9a - 27b + 81c &= 162 \\ 4a - 8b + 16c &= 40 \\ 0 + 0 + 0 &= 0 \\ 9a + 27b + 81c &= 18 \\ 16a + 64b + 256c &= 80 \\ \hline 38a + 56b + 434c &= 300 \end{aligned}$$

The solution of the three normal equations is a simple matter, and we find

$$a = 1.82 \qquad b = -2.65 \qquad c = 0.87$$

The required solution is therefore

$$y = 1.82 - 2.65x + 0.87x^2$$

When, as is often the case, the abscissas of the points to which we wish to fit a polynomial curve are equally spaced, the labor involved in the least-square procedure we have just described can be significantly reduced by using what are known as **orthogonal polynomials.**

DEFINITION 1

If $n + 1$ polynomials $P_{nm}(x)$ of respective degrees $m = 0, 1, 2, \ldots, n$ have the property that

$$(6) \qquad \sum_{x=0}^{n} P_{nj}(x)P_{nk}(x) = 0 \qquad j \neq k$$

they are called orthogonal polynomials.

By methods which need not concern us here,* it has been shown that, for each n, there exists a set of $n + 1$ orthogonal polynomials, and the general formula for them has been obtained

$$(7) \qquad P_{nm}(x) = \sum_{i=0}^{m} (-1)^i \binom{m}{i} \binom{m+i}{i} \frac{(x)^{(i)}}{(n)^{(i)}} \qquad m = 0, 1, 2, \ldots, n$$

In particular,

$$P_{n0}(x) = 1$$

$$P_{n1}(x) = 1 - 2\frac{x}{n}$$

$$P_{n2}(x) = 1 - 6\frac{x}{n} + 6\frac{x(x-1)}{n(n-1)}$$

$$P_{n3}(x) = 1 - 12\frac{x}{n} + 30\frac{x(x-1)}{n(n-1)} - 20\frac{x(x-1)(x-2)}{n(n-1)(n-2)}$$

Clearly, for each $m \leq n$, any polynomial of degree m can be expressed as a linear combination of the polynomials

$$P_{n0}(x),\ P_{n1}(x),\ \ldots,\ P_{nm}(x)$$

---

* See, for instance, W. E. Milne, "Numerical Analysis," pp. 265–275 and 375–381, Princeton University Press, Princeton, N.J., 1949.

for the expression

(8) $$P(x) = a_0P_{n0}(x) + a_1P_{n1}(x) + \cdots + a_mP_{nm}(x)$$

is obviously a polynomial of degree m containing the maximum number, $m + 1$, of independent, arbitrary constants which can appear in the general polynomial of this degree. Moreover, the coefficients $a_0, a_1, \ldots, a_m$ in (8) can easily be found. For, if we multiply both sides of this identity by $P_{ni}(x)$, say, and then sum from $x = 0$ to $x = n$, we get

$$\sum_{x=0}^{n} P(x)P_{ni}(x) = a_0 \sum_{x=0}^{n} P_{n0}(x)P_{ni}(x) + \cdots + a_i \sum_{x=0}^{n} P_{ni}^2(x) + \cdots + a_m \sum_{x=0}^{n} P_{nm}(x)P_{ni}(x)$$

But, from the so-called **orthogonality property** of the polynomials $\{P_{nm}(x)\}$, which is expressed by Eq. (6), it follows that every term on the right-hand side of the last expression is zero except the sum

$$a_i \sum_{x=0}^{n} P_{ni}^2(x)$$

Hence, solving for a_i, we obtain the formula

(9) $$a_i = \frac{\sum\limits_{x=0}^{n} P(x)P_{ni}(x)}{\sum\limits_{x=0}^{n} P_{ni}^2(x)} \qquad i = 0, 1, \ldots, m$$

The property described by (6) is a very important one and we shall encounter it again in Sec. 11.2 when we attempt, in a manner analogous to the expansion (8), to express an arbitrary vector as a linear combination of certain given, independent vectors. Also, in Chaps. 6, 8, and 9 we shall study expansion problems resembling (8), in which the coefficients will be determined through the use of orthogonality properties involving integrals rather than sums, as in (6).

Clearly, an expansion of the form (8) can be created for any function $f(x)$, polynomial or not, merely by using the coefficient formula (9) with $f(x)$ replacing $P(x)$. Such expansions are of great importance, for, although it is obvious that they cannot represent $f(x)$ exactly unless $f(x)$ is a polynomial of degree n or less, they provide the best polynomial approximations to $f(x)$ in the least-square sense. To prove this, suppose that we have a function $f(x)$ defined for the $n + 1$ equally spaced values $x = 0, 1, \ldots, n$, which we wish to approximate with a polynomial of degree m $(<n)$. If we assume the polynomial to be written in the form (8), the discrepancy at the general point x is

$$f(x) - a_0P_{n0}(x) - \cdots - a_iP_{ni}(x) - \cdots - a_mP_{nm}(x)$$

and the principle of least squares requires that we minimize the sum

(10) $$E = \sum_{x=0}^{n} [f(x) - a_0P_{n0}(x) - \cdots - a_iP_{ni}(x) - \cdots - a_mP_{nm}(x)]^2$$

If we equate to zero the derivative of E with respect to a_i, say, we obtain the general minimizing condition

$$\frac{\partial E}{\partial a_i} = \sum_{x=0}^{n} 2[f(x) - a_0P_{n0}(x) - \cdots - a_iP_{ni}(x) - \cdots - a_mP_{nm}(x)]P_{ni}(x) = 0$$

or, breaking up the sum,

(11) $$\sum_{x=0}^{n} f(x)P_{ni}(x) - a_0 \sum_{x=0}^{n} P_{n0}(x)P_{ni}(x) - \cdots - a_i \sum_{x=0}^{n} P_{ni}^2(x) - \cdots - a_m \sum_{x=0}^{n} P_{nm}(x)P_{ni}(x) = 0$$

But, from the orthogonality of the P's, the sums involving two different P's are all zero, and Eq. (11) reduces to

$$\sum_{x=0}^{n} f(x)P_{ni}(x) - a_i \sum_{x=0}^{n} P_{ni}^2(x) = 0$$

or

(12) $$a_i = \frac{\sum_{x=0}^{n} f(x)P_{ni}(x)}{\sum_{x=0}^{n} P_{ni}^2(x)} \qquad i = 0, 1, \ldots, m$$

which is exactly the same as (9) with $P(x)$ replaced by $f(x)$.

The advantage of using orthogonal polynomials is now clear. In the first place, through their use the coefficients in the least-square polynomial approximation to a function $f(x)$ defined for the $n + 1$ equally spaced values $x = 0, 1, \ldots, n$ can be found one at a time without the necessity of solving any simultaneous equations. In the second place, since Formula (12) for a_i does not involve m, the degree of the polynomial we are fitting to the data, it follows that, if we desire to increase m, that is, add another term to the approximating polynomial, all previously calculated coefficients remain unchanged and only the coefficient of the new term need be computed.

The sum appearing in the denominator of (12) need not be calculated directly because a general formula for it is available, namely,

(13) $$\sum_{x=0}^{n} P_{ni}^2(x) = \frac{(n+i+1)^{(i+1)}}{(2i+1)(n)^{(i)}}\dagger$$

---

† See, for instance, Milne, *loc. cit.*

In particular,

$$\sum_{x=0}^{n} P_{n0}^2(x) = n + 1$$

$$\sum_{x=0}^{n} P_{n1}^2(x) = \frac{(n+1)(n+2)}{3n}$$

$$\sum_{x=0}^{n} P_{n2}^2(x) = \frac{(n+1)(n+2)(n+3)}{5n(n-1)}$$

$$\sum_{x=0}^{n} P_{n3}^2(x) = \frac{(n+1)(n+2)(n+3)(n+4)}{7n(n-1)(n-2)}$$

To determine the accuracy with which the polynomial approximation fits the data it is not necessary to compute E from (10), since it can be shown that, in general,

$$E = \sum_{x=0}^{n} f^2(x) - \sum_{i=0}^{m} \left[a_i^2 \sum_{x=0}^{n} P_{ni}^2(x)\right] \tag{14}$$

EXAMPLE 2

Using orthogonal polynomials, fit equations of the form $y = a_0 + a_1 t$ and $y = a_0 + a_1 t + a_2 t^2$ to the data

t	0.00	0.25	0.50	0.75	1.00
y	0.00	0.06	0.20	0.60	0.90

As a first step we must introduce an auxiliary variable $x = 4t$ which will take on the values 0, 1, 2, 3, 4 when t takes on the given values 0.00, 0.25, 0.50, 0.75, 1.00. Then, because there are five given points to which the required curves are to be fitted, we observe that $n + 1 = 5$ or $n = 4$. Next, lacking tables of the orthogonal polynomials, we must compute the values of $P_{40}(x)$, $P_{41}(x)$, and $P_{42}(x)$ for the five values $x = 0, 1, 2, 3, 4$. This is a simple matter, of course, and the values shown in the accompanying table can be calculated at once. It is then necessary to compute the sums of the products of the respective values of the y's and each of the P's. These products are shown in the last three columns of the table.

t	x	y	P_{40}	P_{41}	P_{42}	yP_{40}	yP_{41}	yP_{42}
0.00	0	0.00	1.000	1.000	1.000	0.000	0.000	0.000
0.25	1	0.06	1.000	0.500	−0.500	0.060	0.030	−0.030
0.50	2	0.20	1.000	0.000	−1.000	0.200	0.000	−0.200
0.75	3	0.60	1.000	−0.500	−0.500	0.600	−0.300	−0.300
1.00	4	0.90	1.000	−1.000	1.000	0.900	−0.900	0.900

$$\sum_{x=0}^{4} P_{40}^2 = 5.000 \qquad \sum_{x=0}^{4} yP_{40} = 1.760$$

$$\sum_{x=0}^{4} P_{41}^2 = 2.500 \qquad \sum_{x=0}^{4} yP_{41} = -1.170$$

$$\sum_{x=0}^{4} P_{42}^2 = 3.500 \qquad \sum_{x=0}^{4} yP_{42} = 0.370$$

The coefficients a_0, a_1, and a_2 are then given by Eq. (12):

$$a_0 = \frac{1.760}{5.000} = 0.3520 \qquad a_1 = \frac{-1.170}{2.500} = -0.4680 \qquad a_2 = \frac{0.370}{3.500} = 0.1057$$

In terms of x, the line of best fit is therefore,

$$y = a_0P_{40}(x) + a_1P_{41}(x) = 0.3520 - 0.4680\left(1 - \frac{x}{2}\right) = -0.116 + 0.234x$$

and the parabola of best fit is

$$\begin{aligned} y &= a_0P_{40}(x) + a_1P_{41}(x) + a_2P_{42}(x) \\ &= 0.3520 - 0.4680\left(1 - \frac{x}{2}\right) + 0.1057\left(1 - \frac{3x}{2} + \frac{x^2 - x}{2}\right) \\ &= -0.0103 + 0.0226x + 0.0529x^2 \end{aligned}$$

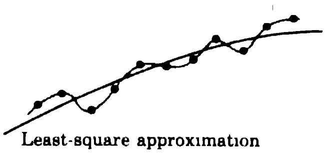
Then, by setting $x = 4t$, we obtain the curves of best fit for the data as originally given

$$y = -0.116 + 0.936t \qquad \text{and} \qquad y = -0.0103 + 0.0904t + 0.8464t^2$$

Using Eq. (14) we find that the sum of the squares of the departures of the points from the line and from the parabola of best fit are, respectively, $E_1 = 0.0465$ and $E_2 = 0.0074$. From the relative size of E_1 and E_2 we conclude that the parabola fits the data significantly better than the straight line does.

In many important applications the position of a moving object is observed at a series of equally spaced times, and its velocity and acceleration are required at these times. Clearly, these can be estimated by using the formulas for numerical differentiation we obtained in Sec. 4.3. However, since the interpolation polynomials from which these formulas of numerical differentiation were derived fit the raw data *exactly*, they and any formulas based on them are seriously influenced by even small errors in the data. On the other hand, a polynomial curve fitted to the data or a portion of the data by the method of least squares will be less influenced by random errors and will represent more nearly the underlying, presumably smooth trend of the data. Hence, derivatives computed from such approximating functions will in general be more accurate than those computed by differentiating interpolation polynomials. These ideas are illustrated geometrically in Fig. 4.5, where it is clear that the

FIGURE 4.5
The relative smoothness of an interpolation polynomial and a least-square approximation to a set of points.



slope of the interpolation polynomial fluctuates markedly from point to point, whereas the slope of the least-square approximation changes in a smooth fashion, which is almost certainly a more reliable description of the trend the data would exhibit if free from random errors

In applying these considerations to the analysis of observed positional data, it is customary to fit a polynomial curve of relatively low degree, say a parabola, to successive sets of 5, 7, or 9 observations and then take the ordinate and the first and second derivatives of this approximating polynomial at the central point of the set as the corrected, or "smoothed," position, velocity, and acceleration of the body at that instant

To illustrate this technique, let us use the method of orthogonal polynomials to fit a parabola to the five points

t	x	y
$-2h$	0	y_{-2}
$-h$	1	y_{-1}
0	2	y_0
h	3	y_1
$2h$	4	y_2

$$x = 2 + \frac{t}{h}$$

To do this, we use the coefficient formula (12) and the values of $P_{40}(x)$, $P_{41}(x)$, and $P_{42}(x)$ tabulated in Example 2. The results are found immediately to be

$$a_0 = \frac{y_{-2} + y_{-1} + y_0 + y_1 + y_2}{5}$$

$$a_1 = \frac{2}{5}\left(y_{-2} + \frac{y_{-1}}{2} - \frac{y_1}{2} - y_2\right)$$

$$a_2 = \frac{2}{7}\left(y_{-2} - \frac{y_{-1}}{2} - y_0 - \frac{y_1}{2} + y_2\right)$$

and thus the formula for the approximating polynomial is

$$\begin{aligned} a_0P_{40}(x) + a_1P_{41}(x) + a_2P_{42}(x) &= \frac{1}{5}(y_{-2} + y_{-1} + y_0 + y_1 + y_2) \\ &\quad + \frac{2}{5}\left(y_{-2} + \frac{y_{-1}}{2} - \frac{y_1}{2} - y_2\right)\left(1 - \frac{x}{2}\right) \\ &\quad + \frac{2}{7}\left(y_{-2} - \frac{y_{-1}}{2} - y_0 - \frac{y_1}{2} + y_2\right)\left(1 - \frac{3}{2}x + \frac{x^2 - x}{2}\right) \end{aligned} \tag{15}$$

When $x = 2$, we find for the smoothed mid-ordinate

$$Y_0 = \frac{-3y_{-2} + 12y_{-1} + 17y_0 + 12y_1 - 3y_2}{35} \tag{16}$$

To approximate dy/dt, we recall that $x = 2 + t/h$; hence

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt} = \frac{1}{h}\frac{dy}{dx}$$

Therefore, differentiating (15) with respect to x and then setting $x = 2$, we find for the smoothed value of the first derivative at the central point of the set

$$Y_0' = \frac{-2y_{-2} - y_{-1} + y_1 + 2y_2}{10h} \tag{17}$$

Similarly, a second differentiation would yield

$$Y_0'' = \frac{2y_{-2} - y_{-1} - 2y_0 - y_1 + 2y_2}{7h^2} \tag{18}$$

as the smoothed value of the second derivative at the central point of the set. However, it is better to find Y_0'' by applying Formula (17) to the table of the smoothed first-derivative values, getting

$$Y_0'' = \frac{-2Y_{-2}' - Y_{-1}' + Y_1' + 2Y_2'}{10h} \tag{19}$$

or, replacing each derivative by its expression in terms of the appropriate y-values from (17),

$$Y_0'' = \frac{4y_{-4} + 4y_{-3} + y_{-2} - 4y_{-1} - 10y_0 - 4y_1 + y_2 + 4y_3 + 4y_4}{100h^2} \tag{20}$$

Since Eqs. (16) and (17) require a knowledge of two ordinates on each side of those being smoothed, it is evident that these equations can be used only for points after the second and before the $(n - 1)$st in a table of data. Similarly, Formula (20) for the second derivative can be used only between the fifth and the $(n - 4)$th points, inclusive. To smooth to the ends of a table we must derive auxiliary formulas from Eq. (15) by evaluating it and its derivatives at $x = 0$, 1, 3, and 4 as well as at $x = 2$. These results will be found among the exercises at the end of this section. In general, central formulas, that is, formulas in which the element being smoothed is as near as possible to the central member of the set of data appearing in the smoothing formula, should be used wherever possible.

The method of least squares is not limited in its application to problems in which the equations of condition are linear. Sometimes, by a suitable transformation, the problem can be converted into one in which the parameters do enter linearly. For instance, to fit an equation of the important type $y = ae^{bx}$, we can take the natural logarithm of each side, getting

$$\ln y = \ln a + bx$$

Then, considering x and $\ln y$ as new variables, say X and Y, and $\ln a$ and b as new parameters, say A and B, we can regard the problem as requiring the determination of A and B such that the *linear* equation

$$Y = A + BX$$

gives the best possible fit to the known pairs of values of X $(= x)$ and Y $(= \ln y)$. Once A has been found, it is, of course, a simple matter to find the actual parameter a, since $A = \ln a$.

Similarly, the fitting of a function $y = kx^n$ can be reduced to a linear problem by first taking logarithms (preferably to the base 10), getting

$$\log y = \log k + n \log x$$

This equation is linear in the parameters $K = \log k$ and $N = n$. Hence the determination of the parameters can be carried out as outlined above.

On the other hand, it is not possible to make a rigorous linearization of general systems of nonlinear equations of condition. But if a reasonable approximation to a solution of such a system is available, an approximate linearization of the problem can be achieved in the following way.

Let the equations to be satisfied (as nearly as possible) be

$$f_1(x,y) = 0, \qquad f_2(x,y) = 0, \qquad \ldots, \qquad f_n(x,y) = 0 \tag{21}$$

and suppose that (x_0,y_0) is known, by inspection or otherwise, to be an approximate solution of this system. Then we can expand each function $f_i(x,y)$ in a generalized Taylor series about the point (x_0,y_0), getting

$$\begin{aligned} f_i(x,y) = f_i(x_0,y_0) &+ \frac{\partial f_i}{\partial x}\bigg|_{x_0,y_0}(x - x_0) + \frac{\partial f_i}{\partial y}\bigg|_{x_0,y_0}(y - y_0) \\ &+ \frac{1}{2}\left[\frac{\partial^2 f_i}{\partial x^2}\bigg|_{x_0,y_0}(x - x_0)^2 + 2\frac{\partial^2 f_i}{\partial x\,\partial y}\bigg|_{x_0,y_0}(x - x_0)(y - y_0)\right. \\ &\qquad\left. + \frac{\partial^2 f_i}{\partial y^2}\bigg|_{x_0,y_0}(y - y_0)^2\right] + \cdots \end{aligned}$$

Now, if (x_0,y_0) is a reasonable approximation to the required solution, the quantities $x - x_0$ and $y - y_0$ will be small, and hence their squares, products, and higher powers will be negligible in comparison with the quantities themselves. Omitting these terms thus reduces the set (21) to the system

$$f_i(x,y) = f_i(x_0,y_0) + \frac{\partial f_i}{\partial x}\bigg|_{x_0,y_0}(x - x_0) + \frac{\partial f_i}{\partial y}\bigg|_{x_0,y_0}(y - y_0) = 0 \tag{22}$$

which is linear in the unknown corrections $x - x_0$ and $y - y_0$. The method of least squares can now be applied to the system (22) in a straightforward way, following which the preliminary estimate (x_0,y_0) can be appropriately corrected. Of course, if desired, the given functions $f_i(x,y)$ can be expanded about the corrected solution (x_1,y_1) and the process repeated. The extension to systems with more than two unknowns

$$f_1(x,y,z,\ldots) = 0, \qquad f_2(x,y,z,\ldots) = 0, \qquad \ldots, \qquad f_n(x,y,z,\ldots) = 0$$

is immediate.

EXAMPLE 3

Fit an equation of the form $y = kx^n$ to the data

x	1	2	3	4
y	2.500	8.000	19.000	50.000

and compute the value of E.

First let us work the problem by using the logarithmic equivalent

$$\log y = \log k + n \log x$$

of the function we are trying to fit to the data. Then the equations of condition are

$$0.3979 = \log k$$
$$0.9031 = \log k + 0.3010n$$
$$1.2788 = \log k + 0.4771n$$
$$1.6990 = \log k + 0.6021n$$

and from these, by the usual process, we obtain the normal equations

$$4.0000 \log k + 1.3802n = 4.2788$$
$$1.3802 \log k + 0.6807n = 1.9049$$

From these we find $\log k = 0.3472$ and $n = 2.096$. Hence, $k = 2.224$, and the required function is

$$y = 2.224x^{2.096}$$

To find E we must evaluate the function $y = 2.224x^{2.096}$ for $x = 1, 2, 3, 4$, subtract these results from the corresponding values of y as originally given, square these differences, and add them. The work is shown in the following table.

x	$y\ (= 2.224x^{2.096})$	y (given)	δ	δ^2
1	2.224	2.500	0.276	0.076
2	9.510	8.000	-1.510	2.280
3	22.243	19.000	-3.243	10.517
4	40.655	50.000	9.345	87.329
				$E = 100.202$

Although we have no real basis for such a conviction, this value of E should strike us as discouragingly large, especially in view of the fact that we have tried to choose the parameters k and n to make it as small as possible. To explore the matter further, let us reconsider the problem in a more elementary way and determine k and n so that the curve will pass exactly through the points (3,19) and (4,50) without regard to the remaining pair of points. This requires that

$$19 = k3^n \qquad \text{and} \qquad 50 = k4^n$$

Dividing the second equation by the first gives us $(\frac{4}{3})^n = \frac{50}{19}$. Hence, taking logs,

$$n = \frac{\log 50 - \log 19}{\log 4 - \log 3} = 3.36$$

With n known, it is easy to find k from the equation $19 = k3^n$

$$\log k = \log 19 - 3.36 \log 3 = 9.67563 \qquad \text{and} \qquad k = 0.474$$

Now, for the function $y = 0.474x^{3.36}$, the calculation of E leads to the following results

x	$y\ (= 0.474x^{3.36})$	y (given)	δ	δ^2
1	0.474	2.500	2.026	4.105
2	4.865	8.000	3.135	9.828
3	19.000	19.000	0.000	0.000
4	50.000	50.000	0.000	0.000
				$E = 13.933$ (!)

This is a remarkable improvement in the closeness of fit, which surely requires explanation

The question will become clearer if we consider the sums of the squares of the errors associated with the respective functions $y = 2.224x^{2.096}$ and $y = 0.474x^{3.36}$ when they are written in logarithmic form These are

x	$\log y\ (= \log 2.224 + 2.096 \log x)$	$\log y$ (given)	δ	δ^2
1	0.3471	0.3979	0.0508	0.00258
2	0.9782	0.9031	−0.0751	0.00564
3	1.3472	1.2788	−0.0684	0.00468
4	1.6091	1.6990	0.0899	0.00808
				$E = 0.02098$

and

x	$\log y\ (= \log 0.474 + 3.36 \log x)$	$\log y$ (given)	δ	δ^2
1	−0.3244	0.3979	0.7233	0.52172
2	0.6871	0.9031	0.2160	0.04666
3	1.2788	1.2788	0.0000	0.00000
4	1.6990	1.6990	0.0000	0.00000
				$E = 0.56838$

The function $y = 2.224x^{2.096}$ which we fitted logarithmically by the method of least squares fits the logarithms of the data much better than does the second function we derived Moreover, it does this by keeping the discrepancies δ_i about equally small However, a given difference δ in the logarithms of two numbers represents only a small difference in the numbers if the logarithms are near zero, but represents a large difference if the logarithms themselves are large Thus, for a change of 0.10000 in the logarithms, we might have either

$$\begin{aligned} 0.10000 &= \text{logarithm of } 1.259 \\ 0.00000 &= \text{logarithm of } 1.000 \\ \text{Difference of the numbers} &= \qquad\qquad 0.259 \end{aligned}$$

or

$$\begin{aligned} 1.60000 &= \text{logarithm of } 39.811 \\ 1.50000 &= \text{logarithm of } 31.623 \\ \text{Difference of the numbers} &= \qquad\qquad 8.188 \end{aligned}$$

Hence, the average approximation to the original data is significantly improved by keeping the errors in the larger logarithms as small as possible, even at the expense of considerably larger errors in the smaller logarithms. And, clearly, there is no reason to believe that the function which best fits the logarithms of the data will necessarily give the best approximation to the data themselves.

As a final approach to the problem, let us now try the general method of handling nonlinear equations of condition. Assuming again an equation of the form $y = kx^n$ and substituting the four given sets of values, we find that k and n should satisfy the conditions

$$\begin{aligned} 2.5 &= k \\ 8.0 &= k2^n \\ 19.0 &= k3^n \\ 50.0 &= k4^n \end{aligned}$$

As an initial estimate of the values of k and n, let us use the values $k = 0.474$ and $n = 3.36$ which we obtained by passing the curve exactly through the points (3,19) and (4,50). Then expanding each of the equations of condition in a Taylor series around (0.474,3.36), we find

$$\begin{aligned} f_1 &\equiv k - 2.500 = -2.026 + (k - 0.474) = 0 \\ f_2 &\equiv k2^n - 8.000 = (4.865 - 8.000) + 2^n\Big|_{0.474,3.36}(k - 0.474) \\ &\qquad + k2^n \ln 2\Big|_{0.474,3.36}(n - 3.36) \\ &= -3.135 + 10.267(k - 0.474) + 3.372(n - 3.36) = 0 \\ f_3 &\equiv k3^n - 19.000 \doteq (19.000 - 19.000) + 3^n\Big|_{0.474,3.36}(k - 0.474) \\ &\qquad + k3^n \ln 3\Big|_{0.474,3.36}(n - 3.36) \\ &= 40.098(k - 0.474) + 20.874(n - 3.36) = 0 \\ f_4 &\equiv k4^n - 50.000 \doteq (50.000 - 50.000) + 4^n\Big|_{0.474,3.36}(k - 0.474) \\ &\qquad + k4^n \ln 4\Big|_{0.474,3.36}(n - 3.36) \\ &= 105.411(k - 0.474) + 69.314(n - 3.36) = 0 \end{aligned}$$

Letting $u = k - 0.474$ and $v = n - 3.36$, the approximate equations of condition are, therefore,

$$\begin{aligned} u \qquad\qquad &= 2.026 \\ 10.267u + 3.372v &= 3.135 \\ 40.098u + 20.874v &= 0.000 \\ 105.411u + 69.314v &= 0.000 \end{aligned}$$

The construction of the normal equations, by multiplying each equation of condition first by the coefficient of u and then by the coefficient of v in that equation and adding, is a routine matter, and we find without difficulty

$$\begin{aligned} 12{,}825.740u + 8{,}178.084v &= 34.213 \\ 8{,}178.084u + 5{,}251.525v &= 10.571 \end{aligned}$$

Hence, $\quad u = 0.197 \quad$ and $\quad v = -0.305$

and the corrected estimates of k and n are

$$k = 0.474 + 0.197 = 0.671$$

$$n = 3.36 - 0.305 = 3.055$$

For the function $y = 0.671x^{3.055}$ a straightforward calculation yields $E = 22.628$, which is still not so small as the value we found for the curve that passed exactly through the points (3,19) and (4,50). However, a second application, based upon expanding the equations of condition around $k = 0.671$ and $n = 3.055$, yields the improved values

$$k = 0.733 \qquad \text{and} \qquad n = 3.039$$

and $E = 10.052$, which is the smallest value of E we have yet found. Another repetition of the process would no doubt improve this slightly.

EXERCISES

1. Fit a straight line to the data

x	1	3	6	7	9
y	1	5	6	10	12

(a) by minimizing the sum of the squares of the vertical distances from the points to the line, and (b) by minimizing the sum of the squares of the horizontal distances from the points to the line.

2. Fit an equation of the form $y = a + bx + cx^2$ to the data

x	-1	0	2	3	5
y	-4	4	8	9	7

3. Find the most plausible values of x and y from the following system of equations

$$x + y = 2$$
$$2x - 3y = 9$$
$$20x + 16y = 4$$

(a) without dividing out the factor 4 from the last equation, and (b) after dividing out the factor 4 from the last equation. Explain.

4. Fit equations of each of the forms

a $ax + by - 1 = 0$ **b** $ax + y - c = 0$ **c** $x + by - c = 0$

to the data

x	0	1	2	3
y	1.1	1.9	3.0	3.9

by minimizing the sum of the squares of the amounts by which each of the equations, in turn, fails to be satisfied. Compare the results and explain the differences.

5 **a** Verify that

$$\sum_{x=0}^{n} P_{n0}(x)P_{n1}(x) = 0 \qquad \sum_{x=0}^{n} P_{n0}(x)P_{n2}(x) = 0 \qquad \sum_{x=0}^{n} P_{n1}(x)P_{n2}(x) = 0$$

b Verify that

$$\sum_{x=0}^{n} P_{n0}^2(x) = n + 1 \qquad \sum_{x=0}^{n} P_{n1}^2(x) = \frac{(n+1)(n+2)}{3n}$$

$$\sum_{x=0}^{n} P_{n2}^2(x) = \frac{(n+1)(n+2)(n+3)}{5n(n-1)}$$

c Express x and x^2 as series of the form

$$a_0P_{n0}(x) + a_1P_{n1}(x) + a_2P_{n2}(x) +$$

6 Using orthogonal polynomials, fit functions of each of the forms $y = a + bx$ and $y = a + bx + cx^2$ to the data

x	0.50	1.00	1.50	2.00	2.50	3.00
y	1.01	1.08	1.16	1.25	1.29	1.30

and compute the value of E for each approximation

7 Fit an equation of the form $y = Ae^{ax}$ to the data

x	1	2	3	4
y	1.65	2.70	4.50	7.35

a By first taking logarithms and then working with the linearized equation

$$\ln y = \ln A + ax$$

b By first obtaining approximate values of A and a and then linearizing by expanding the equations of condition in Taylor's series around these values and retaining only the linear terms.

8 Derive the following modifications of Formulas (16) and (17)

$$Y_0 = \tfrac{1}{35}(31y_0 + 9y_1 - 3y_2 - 5y_3 + 3y_4)$$

$$Y_0' = \frac{1}{70h}(-54y_0 + 13y_1 + 40y_2 + 27y_3 - 26y_4)$$

and

$$Y_0 = \tfrac{1}{35}(9y_{-1} + 13y_0 + 12y_1 + 6y_2 - 5y_3)$$

$$Y_0' = \frac{1}{70h}(-34y_{-1} + 3y_0 + 20y_1 + 17y_2 - 6y_3)$$

9 Derive Formula 14

10 It is desired to fit a circular arc to a set of points (x_1,y_1), , (x_n,y_n) Discuss the relative merits of doing this by minimizing the sum of the squares of the vertical distances from the points to the circular arc and by taking the equation of the circle in the form $x^2 + y^2 +$

$ax + by + c = 0$ and minimizing the sum of the squares of the amounts by which the coordinates of the points fail to satisfy this equation

11 It is desired to fit an equation of the form $y = Ae^{ax}$ to a set of points $(x_1,y_1), \ldots, (x_n,y_n)$ By observing that y must satisfy a certain linear, constant-coefficient, first-order difference equation, obtain the following equations of condition

$$\begin{aligned} y_2 - e^a y_1 &= 0 \\ y_3 - e^a y_2 &= 0 \\ &\vdots \\ y_n - e^a y_{n-1} &= 0 \end{aligned}$$

Show how A can be found after the best least-square approximation to a has been found from these equations Discuss the advantages of this method relative to the method of linearizing by taking logarithms and the general method for handling problems in which the parameters enter nonlinearly

12 Explain how the method of Exercise 11 can be extended to the fitting of functions of the form $y = Ae^{ax} + Be^{bx} + \cdots + Ke^{kx}$

13 Explain how a continuous function can be approximated over an interval (a,b) by minimizing the integral of the squared difference between the given function and the chosen approximation Illustrate by approximating the function $\cos x$ over the interval $(0,\pi/2)$ with a function of the form $y = a - bx^2$

14 Approximate the solution of $y'' + x^2 y = \sin x$ for which $y(0) = y(\pi) = 0$ by assuming $y = A \sin x$ and choosing to minimize the integral from 0 to π of the square of the amount by which $A \sin x$ fails to satisfy the differential equation

15 Show that if a line with equation $x \cos \theta + y \sin \theta - p = 0$ is fitted to a set of points $(x_1,y_1), \ldots, (x_n,y_n)$ by minimizing the sum of the squares of the perpendicular distances from the points to the line, the value of θ is given by the formula

$$\tan 2\theta = \frac{2r_{xy}\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}$$

where σ_x and σ_y are, respectively, the so-called **standard deviations** of the x-values and the y-values,

$$\sigma_x = \frac{1}{n}\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \quad \text{and} \quad \sigma_y = \frac{1}{n}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}$$

and r_{xy} is the **coefficient of correlation** between the x-values and the y-values,

$$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n^2\sigma_x\sigma_y}$$

What is the value of p in the equation of the line of best fit?

CHAPTER FIVE

Mechanical and Electrical Circuits

5.1

Introduction

An examination of the application of differential equations to mechanical and electrical systems is valuable for at least two reasons. In the first place, it will furnish us with useful information about the behavior of certain physical systems of great practical interest Second, and perhaps more important, it will provide a striking example of the role which mathematics plays in unifying widely differing phenomena For instance, we shall see that, merely by a renaming of the variables, the analysis of the motion of a weight vibrating on a spring becomes the analysis of a simple electrical circuit Moreover, this correspondence is not merely qualitative or descriptive It is quantitative, in the sense that if one is given any of a wide variety of vibrating mechanical systems, an electrical circuit can be constructed whose currents or voltages, as preferred, will give the *exact* values of the displacements in the mechanical system when suitable scale factors are introduced Since electrical circuits are easy to assemble and since currents and voltages are easy to measure, this affords a practical method of studying the vibration of complicated mechanical configurations, such as engine crankshafts, which are expensive to make and modify and whose motions are difficult to record accurately *

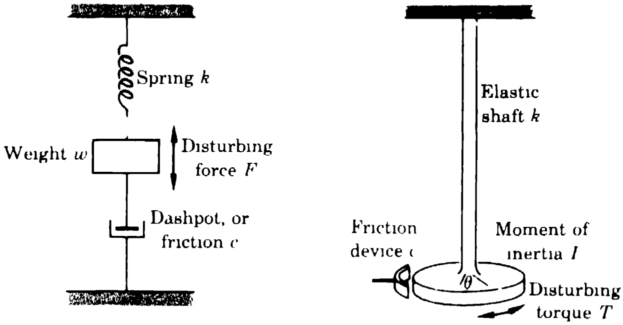
5.2

Systems with one degree of freedom

A system which can be described completely by one coordinate, i e , by one physical datum such as a displacement, an angle, a current, or a voltage, is called **a system of one degree of freedom.** A system requiring more than one coordinate for its complete description is called **a system of several degrees of freedom.** A single differential equation suffices for the mathematical description of a system of one degree of freedom A set of simultaneous differential equations, as many equations as there are degrees of freedom, is necessary for the analysis of systems of more than one degree of freedom. We shall begin our investigations by considering, as prototypes of the general system with one degree of freedom, each of the configurations shown in Fig. 5 1 In each case we assume that all the elements of the system are concentrated, or **lumped.** In other words, such things as the distributed mass of the spring in Fig 5 1*a*, the distributed moment of inertia of the shaft in Fig. 5 1*b*, and the resistance of the leads in Fig 5 1*c* and *d* we assume to be either negligible or taken into account through suitable corrections added to the corresponding major elements *

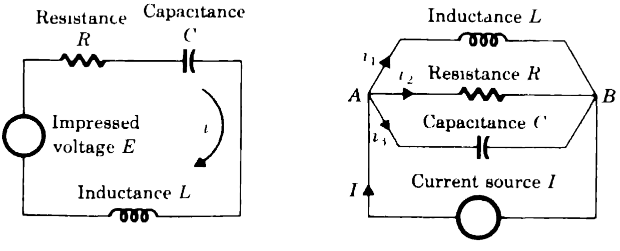
---

* Of course, mechanical models of electrical circuits can also be constructed, but there is little practical reason for constructing them



(*a*) Coordinate = vertical displacement of weight y

(*b*) Coordinate – angular displacement of disk θ



(*c*) Coordinate = current i, flowing around loop

(*d*) Coordinate = common voltage e, between nodes A and B

FIGURE 5 1 Four simple systems of one degree of freedom (*a*) translational-mechanical (*b*) torsional-mechanical, (*c*) series-electrical, (*d*) parallel-electrical

---

* In many problems these assumptions are not sufficiently accurate, and the continuous distribution of the components of the system must be considered As we shall see in Chap 8, this leads to *partial* rather than *ordinary* differential equations

In Fig. 5 1a we assume the weight to be guided, so that only vertical motion, without swinging, is possible As indicated, the effect of friction is not neglected. Instead, we suppose that a retarding force proportional to the velocity acts at all times. Friction of this sort is known as **viscous friction** or **viscous damping.** Its existence is well established for moderate velocities, although for large velocities the resistance may be more nearly proportional to the square or even the cube of the velocity

The analysis of this system is based upon Newton's law,

(1) Mass × acceleration = force

Measuring the displacement y from the equilibrium position of the weight, with the positive direction upward, we have

(2) $$\text{Acceleration of the mass} = \frac{d^2y}{dt^2}$$

The most obvious force acting on the mass is the attraction of gravity·

(3) $$\text{Gravitational force} = -w$$

the minus sign indicating that this force acts downward To compute the elastic force, we observe first that a weight w when hung on a spring of modulus k, that is, a spring requiring k units of force to extend it one unit of length, will stretch the spring a distance equal to w/k Hence, when the weight moves from this equilibrium level during the course of its motion, the instantaneous elongation of the spring is $w/k - y$ If this quantity is positive, the spring is stretched and, therefore, applies to the mass a force which acts in the upward, or positive, direction If this quantity is negative, the spring is compressed and, therefore, applies to the mass a force which acts in the downward, or negative, direction The force the spring exerts on the mass at any time is, therefore,

Force per unit elongation × instantaneous elongation

or

(4) $$\text{Elastic force} = k\left(\frac{w}{k} - y\right) = w - ky$$

To determine the frictional force, we observe that the velocity of the mass is dy/dt, hence, from the assumption of viscous damping,

(5) $$\text{Frictional force} = -c\,\frac{dy}{dt}$$

the minus sign indicating that the resistance always acts in opposition to the velocity Finally, through some external agency, a disturbing force, usually periodic, may act upon the system, upsetting its condition of equilibrium We shall consider specifically the important case in which

(6) $$\text{Impressed force} = F_0 \cos \omega t \qquad F_0 \text{ a constant}$$

Substituting from Eqs (2) to (6) into Newton's law, Eq. (1), we thus have

$$\frac{w}{g}\frac{d^2y}{dt^2} = -w + (w - ky) - c\frac{dy}{dt} + F_0 \cos \omega t$$

or

(7) $$\frac{w}{g}\frac{d^2y}{dt^2} + c\frac{dy}{dt} + ky = F_0 \cos \omega t$$

We note from this equation that the gravitational force on the weight is canceled by that part of the elastic force due to the initial elongation of the spring Because of this, in the future we shall neglect gravitational forces from the outset in the analysis of problems of this sort

Equation (7) is a typical nonhomogeneous, linear differential equation of the second order with constant coefficients, whose general solution we can easily find by the methods of Chap 2 Presumably it will be accompanied by given initial conditions

$$y(0) = y_0 \qquad \text{and} \qquad \left.\frac{dy}{dt}\right|_{t=0} = y_0'$$

and, by using these, the constants in any complete solution can be determined. However, before continuing with the solution of Eq (7) we shall derive the equations governing the other systems shown in Fig 5 1

The analysis of the system of Fig 5 1*b* is based upon Newton's law in torsional form

(8) Moment of inertia × angular acceleration = torque

In this case the various torques are

(9) Elastic torque due to the twisting of the shaft $= -k\theta$

(10) Viscous damping torque $= -c\dfrac{d\theta}{dt}$

(11) Impressed torque $= T_0 \cos \omega t$

Since the angular acceleration is $d^2\theta/dt^2$, on substituting into Newton's law, Eq (8), we have

$$I\frac{d^2\theta}{dt^2} = -k\theta - c\frac{d\theta}{dt} + T_0 \cos \omega t$$

or

(12) $$I\frac{d^2\theta}{dt^2} + c\frac{d\theta}{dt} + k\theta = T_0 \cos \omega t$$

This, too, is a completely familiar differential equation, and, when accompanied by the initial conditions

$$\theta(0) = \theta_0 \qquad \text{and} \qquad \left.\frac{d\theta}{dt}\right|_{t=0} = \theta_0'$$

it can easily be solved for the function describing the behavior of any particular system

The analysis of the series, or one-loop, electrical circuit shown in Fig. 5.1c is based on **Kirchhoff's second law:** *The algebraic sum of the potential differences around any closed loop is zero*, or *the voltage impressed on a closed loop is equal to the sum of the voltage drops in the rest of the loop* Using well-known electrical laws, we have

(13) Voltage drop across the resistance $= iR$

(14) Voltage drop across the condenser $= \frac{1}{C}\int^t i\,dt$

(15) Voltage drop across the inductance $= L\frac{di}{dt}$

Thus, assuming the important case in which

(16) Impressed voltage $= E_0 \cos \omega t$

on substituting from Eqs. (13) to (16) into Kirchhoff's second law, we have

(17) $$L\frac{di}{dt} + iR + \frac{1}{C}\int^t i\,dt = E_0 \cos \omega t$$

Strictly speaking, this is not a differential equation, but rather an **integrodifferential equation.** The operational methods we shall develop in Chap. 7 will handle it directly, but, before we can apply the techniques we have available at this stage, we must convert it into a pure differential equation. There are two ways of doing this. The first is to regard not i but $\int^t i\,dt$ as the dependent variable of the problem. This is not merely a mathematical strategem, for the quantity

$$Q = \int^t i\,dt$$

that is, the integrated flow of current into the condenser, is precisely the quantity of electricity, or electric charge, instantaneously present on the condenser. In terms of Q, then, we have the equation

(18a) $$L\frac{d^2Q}{dt^2} + R\frac{dQ}{dt} + \frac{1}{C}Q = E_0 \cos \omega t$$

subject, of course, to the given initial conditions

$$Q(0) = \int^{t=0} i\,dt = Q_0 \qquad \text{and} \qquad \left.\frac{dQ}{dt}\right|_{t=0} = i(0) = i_0$$

On the other hand, we can also convert Eq. (17) into a differential equation simply by differentiating it with respect to time, getting

(18b) $$L\frac{d^2i}{dt^2} + R\frac{di}{dt} + \frac{1}{C}i = -\omega E_0 \sin \omega t$$

The initial conditions required for an equation of this form are

$$i(0) = i_0 \qquad \text{and} \qquad \left.\frac{di}{dt}\right|_{t=0} = i_0'$$

The first of these was given for the original equation The second can be found from the original equation, since

$$\frac{di}{dt} = \frac{1}{L}\left(E_0 \cos \omega t - iR - \frac{1}{C}\int^t i\,dt\right)$$

and the right-hand side is completely known at $t = 0$

To establish the differential equation describing the behavior of the parallel, or one-node-pair, electrical circuit shown in Fig 5 1*d*, we must use **Kirchhoff's first law**: *The algebraic sum of the currents flowing toward any point in an electrical circuit is zero* Solving for i in Eqs (13), (14), and (15) we obtain, respectively,

(19) Current through the resistance $= \dfrac{e}{R}$

(20) Current (apparently) through the condenser $= C\dfrac{de}{dt}$

(21) Current through the inductance $= \dfrac{1}{L}\displaystyle\int^t e\,dt$

Thus, assuming the important case of a current source such that

(22) Impressed current $= I_0 \cos \omega t$

on substituting from Eqs (19) to (22) into Kirchhoff's first law, we have

$$C\frac{de}{dt} + \frac{1}{R}e + \frac{1}{L}\int^t e\,dt = I_0 \cos \omega t \tag{23}$$

Again, our derivation has led to an integrodifferential equation To convert it to a pure differential equation we can consider $\int^t e\,dt = U$, say, as a new variable, getting

$$C\frac{d^2U}{dt^2} + \frac{1}{R}\frac{dU}{dt} + \frac{1}{L}U = I_0 \cos \omega t \tag{24a}$$

subject to initial conditions of the form

$$U(0) = \int^{t=0} e\,dt = U_0 \qquad \text{and} \qquad \left.\frac{dU}{dt}\right|_{t=0} = e_0$$

On the other hand, we can simply differentiate Eq. (23) with respect to time, getting

$$C\frac{d^2e}{dt^2} + \frac{1}{R}\frac{de}{dt} + \frac{1}{L}e = -\omega I_0 \sin \omega t \tag{24b}$$

subject to initial conditions of the form

$$e(0) = e_0 \qquad \text{and} \qquad \left.\frac{de}{dt}\right|_{t=0} = e_0'$$

When we collect the differential equations we have derived,

$$\text{(7)}\qquad \frac{w}{g}\frac{d^2y}{dt^2} + c\frac{dy}{dt} + ky = F_0 \cos \omega t \qquad \text{(translational-mechanical)}$$

$$\text{(12)}\qquad I\frac{d^2\theta}{dt^2} + c\frac{d\theta}{dt} + k\theta = T_0 \cos \omega t \qquad \text{(torsional-mechanical)}$$

$$\left.\begin{aligned} &\text{(18a)}\qquad L\frac{d^2Q}{dt^2} + R\frac{dQ}{dt} + \frac{Q}{C} = E_0 \cos \omega t \\ &\text{(18b)}\qquad L\frac{d^2i}{dt^2} + R\frac{di}{dt} + \frac{i}{C} = -\omega E_0 \sin \omega t \end{aligned}\right\} \text{(series-electrical)}$$

$$\left.\begin{aligned} &\text{(24a)}\qquad C\frac{d^2U}{dt^2} + \frac{1}{R}\frac{dU}{dt} + \frac{U}{L} = I_0 \cos \omega t \\ &\text{(24b)}\qquad C\frac{d^2e}{dt^2} + \frac{1}{R}\frac{de}{dt} + \frac{e}{L} = -\omega I_0 \sin \omega t \end{aligned}\right\} \text{(parallel-electrical)}$$

their essential mathematical identity becomes apparent. Moreover, we can see the possibility of various physical analogies. For instance, if we compare the translational-mechanical and the series-electrical systems, we find that

Mass $\frac{w}{g} \leftrightarrow$ inductance L

Friction $c \leftrightarrow$ resistance R

Spring modulus $k \leftrightarrow$ elastance $\frac{1}{C}$

Impressed force $F \leftrightarrow \begin{cases} \text{impressed voltage } E \text{ [using Eq. (18}a\text{)]} \\ dE/dt \text{ [using Eq. (18}b\text{)]} \end{cases}$

Displacement $y \leftrightarrow \begin{cases} \text{charge } Q \text{ [using Eq. (18}a\text{)]} \\ \text{current } i \text{ [using Eq. (18}b\text{)]} \end{cases}$

and, if we compare the translational-mechanical and the parallel-electrical systems, we have the correspondences

Mass $\frac{w}{g} \leftrightarrow$ capacitance C

Friction $c \leftrightarrow$ conductance $\frac{1}{R}$

Spring modulus $k \leftrightarrow$ susceptance $\frac{1}{L}$

Impressed force $F \leftrightarrow \begin{cases} \text{impressed current } I \text{ [using Eq. (24}a\text{)]} \\ dI/dt \text{ [using Eq. (24}b\text{)]} \end{cases}$

Displacement $y \leftrightarrow \begin{cases} \int^t e\,dt \text{ [using Eq. (24}a\text{)]} \\ \text{voltage } e \text{ [using Eq. (24}b\text{)]} \end{cases}$

We shall not pursue these analogies further. Instead we shall investigate one or two of the systems in detail.

5.3

The translational-mechanical system

The displacement y of the weight w in the translational-mechanical system (Fig. 5.1a) has been shown to satisfy the differential equation

(1) $$\frac{w}{g}\frac{d^2y}{dt^2} + c\frac{dy}{dt} + ky = F_0 \cos \omega t$$

Following the general theory of Chap. 2, it must therefore consist of two parts. The *complementary function*, obtained by solving Eq. (1) when the term representing the impressed force is deleted, describes the motion of the weight in the absence of any external disturbance. This intrinsic, or natural, behavior of the system is called the **free motion.** The *particular integral* describes the response of the system to a specific influence external to the system. The behavior it represents is called the **forced motion.**

The nature of the free motion of the system will depend upon the roots of the characteristic equation

$$\frac{w}{g}m^2 + cm + k = 0$$

namely, $$m = -\frac{cg}{2w} \pm \frac{g}{2w}\sqrt{c^2 - \frac{4kw}{g}}$$

Since g, w, and k are all positive and c is nonnegative, and since the radical, when real, is certainly less than c, it follows that the real parts of the roots m_1 and m_2 are always negative. We must now consider three possibilities

$$c^2 - \frac{4kw}{g}\begin{cases} > 0 \\ = 0 \\ < 0 \end{cases}$$

In the first case, $c^2 - 4kw/g > 0$, there is a relatively large amount of friction, and, naturally enough, the system is said to be **overdamped.** The free motion, i.e., the motion described by the complementary function, is now given by the expression

$$y = Ae^{m_1t} + Be^{m_2t}$$

where, as we pointed out above, both m_1 and m_2 are negative. Thus y approaches zero as time increases indefinitely. This, of course, is perfectly consistent with the familiar observation that if a system upon which no external forces are acting is displaced from its equilibrium position, it will eventually return to that position as friction causes the disturbance to subside.

If we set $y = 0$, we obtain the equation

$$Ae^{m_1t} + Be^{m_2t} = 0 \qquad \text{or} \qquad e^{(m_1-m_2)t} = -\frac{B}{A} \qquad A \neq 0$$

If A and B, which will, of course, be determined by the initial conditions of the problem, are of opposite sign, then there is one and only one value of t which satisfies the last equation On the other hand, since a real exponential function must always be positive, it follows that when A and B have the same sign or when one or the other of them is zero, there is no time when $y = 0$ A plot of the displacement y during the free motion of an overdamped system must therefore resemble one of the curves shown in Fig 5 2 or the reflection of one of these curves in the t-axis Figure 5 2a, b, and c illustrates the possibilities when A and B are of opposite sign and y vanishes once and only once Assuming that the weight starts its motion when $t = 0$, the zero of y may, of course, occur in the physically irrelevant interval $-\infty < t < 0$ Figure 5 2d illustrates both the case when A and B are of like sign and the case when either A or B is zero and y can never vanish

If $$c^2 - \frac{4kw}{g} = 0$$

we have the border-line case in which the roots of the characteristic equation are real and equal

$$m_1 = m_2 = -\frac{cg}{2w}$$

When this occurs, the motion is said to be **critically damped,** and the exact value of the damping which produces it, namely,

(2) $$c_c = 2\sqrt{\frac{kw}{g}}$$

is known as the **critical damping.** In this case the free motion is given by

$$y = Ae^{m_1 t} + Bte^{m_1 t}$$

If we set $y = 0$, we obtain

$$Ae^{m_1 t} + Bte^{m_1 t} = 0 \qquad \text{or} \qquad t = -\frac{A}{B} \qquad B \neq 0$$

If $B = 0$, there is no value of t for which $y = 0$, but in all other cases there is one and only one value of t for which $y = 0$ This may be in the physically irrelevant interval $-\infty < t < 0$, however, and so it is possible that y will not vanish in the actual motion even when $B \neq 0$ Clearly, there is no essential difference

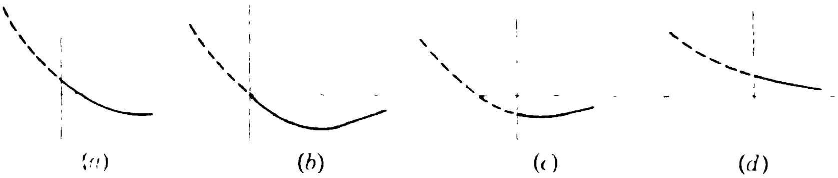

(a) (b) (c) (d)

FIGURE 5 2
Displacement-time plots for free overdamped and critically damped motion

in the character of the motion in the overdamped and critically damped cases, and the possible plots of the displacement y in the critically damped case are also represented by the curves of Fig. 5.2 and their reflections in the t-axis.

If $c^2 - (4kw/g) < 0$, the motion is said to be **underdamped.** The roots of the characteristic equation in this case are the conjugate complex numbers

$$m_1, m_2 = -\frac{cg}{2w} \pm i\frac{g}{2w}\sqrt{\frac{4kw}{g} - c^2} = -p \pm iq$$

where

$$p = \frac{cg}{2w} \qquad \text{and} \qquad q = \frac{g}{2w}\sqrt{\frac{4kw}{g} - c^2} \tag{3}$$

The free motion is therefore described by

$$y = e^{-pt}(A\cos qt + B\sin qt) \tag{4a}$$

or equally well by

$$y = Ge^{-pt}\cos(qt - H) \tag{4b}$$

or by

$$y = Ke^{-pt}\sin(qt - L) \tag{4c}$$

where A, B, G, H, K, and L are arbitrary constants.

The motion described by either (4a), (4b), or (4c) is known as a **damped oscillation,** and its general appearance is shown in Fig. 5.3. It is not periodic, since the factor multiplying the trigonometric terms is continuously decreasing. However, there are regularly spaced passages through the equilibrium position at intervals of π/q. In fact, using the description of the motion provided by Eq. (4b), it is clear that $y = 0$ whenever

$$\cos(qt - H) = 0$$

that is, when $qt - H = \pi/2 + n\pi$ or

$$t = \frac{1}{q}\left(H + \frac{\pi}{2}\right) + \frac{n\pi}{q} \qquad n = 0, 1, 2, \ldots$$

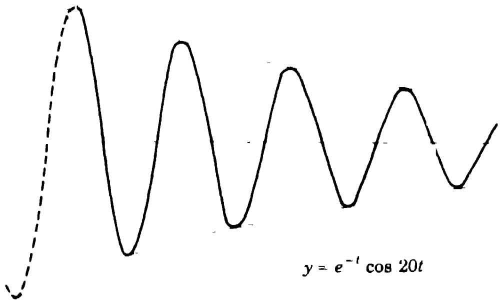


FIGURE 5.3 A typical displacement-time plot for an underdamped system

Hence, we can speak of the *pseudo period* $2\pi/q$ and of the *pseudo frequency* or *"frequency with damping"* ω_d, defined by the equation

$$\frac{\omega_d}{2\pi} = \frac{q}{2\pi} = \frac{1}{2\pi}\,\frac{g}{2w}\sqrt{\frac{4kw}{g} - c^2} = \frac{1}{2\pi}\sqrt{\frac{kg}{w} - \frac{c^2g^2}{4w^2}} \qquad \text{"cycles"/unit time} \tag{5}$$

If $c = 0$, that is, if there is no damping in the system, the motion is strictly periodic, and its frequency, which we shall call the **undamped natural frequency** ω_n, is, from (5), given by the formula

$$\frac{\omega_n}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{kg}{w}} \qquad \text{cycles/unit time} \tag{6}$$

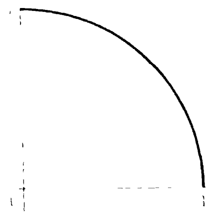
Clearly the "frequency" when damping is present is always less than the undamped natural frequency The ratio of the two frequencies is

$$\frac{\omega_d}{\omega_n} = \frac{\sqrt{kg/w - c^2g^2/4w^2}}{\sqrt{kg/w}}$$

$$= \sqrt{1 - \frac{c^2g}{4kw}} = \sqrt{1 - \frac{c^2}{c_c^2}}$$

since, from Eq (2), $c_c^2 = 4kw/g$ Figure 5 4 shows a plot of ω_d/ω_n versus c/c_c Evidently, if the actual damping is only a small fraction of the critical damping, as it often is, its effect upon the frequency of the motion is very small This explains why friction is usually neglected in natural-frequency calculations

FIGURE 5 4
Plot showing the effect of friction on frequency in an underdamped system



Still using Eq. (4*b*), it is clear that the extreme values of y occur when

$$\frac{dy}{dt} = G\,[-\,pe^{-pt}\cos(qt - H) - qe^{-pt}\sin(qt - H)] = 0$$

that is, when $\tan(qt - H) = -p/q$, or, finally, when

$$t = \frac{H}{q} - \frac{1}{q}\,\mathrm{Tan}^{-1}\frac{p}{q} + \frac{n\pi}{q} = T + \frac{n\pi}{q}$$

where T denotes the constant $(H/q) - (1/q)\,\mathrm{Tan}^{-1}(p/q)$

The ratio of successive extreme displacements on the same side of the equilibrium position is a quantity of considerable importance Its value is

$$\frac{y_n}{y_{n+2}} = \frac{y(T + n\pi/q)}{y[T + (n+2)\pi/q]} = \frac{Ge^{-p(T+n\pi/q)} \cos\left[q\left(T + \frac{n}{q}\pi\right) - H\right]}{Ge^{-p[T+(n+2)\pi/q]} \cos\left[q\left(T + \frac{n+2}{q}\pi\right) - H\right]}$$

$$= e^{2\pi p/q} \frac{\cos(qT + n\pi - H)}{\cos(qT + n\pi - H + 2\pi)}$$

$$(7) \qquad = e^{2\pi p/q}$$

Since this result depends only on the parameters of the system and not on n, we have thus established the following remarkable result

THEOREM 1

The ratio of successive maximum (or minimum) displacements remains constant throughout the entire free motion of an underdamped system

If we take the natural logarithm of Expression (7), we have

$$(8) \qquad \ln \frac{y_n}{y_{n+2}} = \frac{2\pi p}{q}$$

This quantity is known as the **logarithmic decrement** δ, and it is a convenient measure, in **nepers per cycle**, of the rate at which the motion dies away * Substituting for p and q from (3) into (8), we find

$$\delta = \frac{2\pi p}{q} = 2\pi \frac{cg/2w}{(g/2w)\sqrt{(4kw/g) - c^2}} = 2\pi \frac{c}{\sqrt{c_c^2 - c^2}}$$

Solved for c/c_c, this becomes

$$(9) \qquad \frac{c}{c_c} = \frac{\delta}{\sqrt{\delta^2 + 4\pi^2}}$$

Since y_n and y_{n+2} are quantities relatively easy to measure, δ can easily be computed Then from Eq (9) the fraction of critical damping present in a given system can be found at once

Now that we have investigated the free motion of the translational-mechanical system in the overdamped, critically damped, and underdamped cases, it remains for us to consider the forced motion To do this we must, of course, find a particular integral for Eq (1)

$$(1) \qquad \frac{w}{g}\frac{d^2y}{dt^2} + c\frac{dy}{dt} + ky = F_0 \cos \omega t$$

Assuming, as usual,

$$Y = A \cos \omega t + B \sin \omega t$$

---

* Equivalently, though less conventionally, the rate of attenuation could be expressed in **decibels per cycle** by means of the definition

$$\text{Decibels} = 20 \log \frac{y_n}{y_{n+2}}$$

and substituting into (1), collecting terms, and equating to zero the coefficients of $\cos \omega t$ and $\sin \omega t$, we obtain the two conditions

$$\left(k - \omega^2 \frac{w}{g}\right) A + \omega c B = F_0$$

$$-\omega c A + \left(k - \omega^2 \frac{w}{g}\right) B = 0$$

from which we find immediately

$$A = \frac{k - \omega^2(w/g)}{[k - \omega^2(w/g)]^2 + (\omega c)^2} F_0$$

$$B = \frac{\omega c}{[k - \omega^2(w/g)]^2 + (\omega c)^2} F_0$$

Hence,

$$\begin{aligned} Y &= F_0 \frac{[k - \omega^2(w/g)] \cos \omega t + (\omega c) \sin \omega t}{[k - \omega^2(w/g)]^2 + (\omega c)^2} \\ &= \frac{F_0}{\sqrt{[k - \omega^2(w/g)]^2 + (\omega c)^2}} \left\{ \frac{k - \omega^2(w/g)}{\sqrt{[k - \omega^2(w/g)]^2 + (\omega c)^2}} \cos \omega t \right. \\ &\qquad \left. + \frac{\omega c}{\sqrt{[k - \omega^2(w/g)]^2 + (\omega c)^2}} \sin \omega t \right\} \end{aligned}$$

Now, referring to the triangle shown in Fig. 5.5, it is evident that Y can be written in either of the equivalent forms

$$Y = \frac{F_0}{\sqrt{[k - \omega^2(w/g)]^2 + (\omega c)^2}} (\cos \omega t \cos \alpha + \sin \omega t \sin \alpha)$$

$$= \frac{F_0}{\sqrt{[k - \omega^2(w/g)]^2 + (\omega c)^2}} \cos (\omega t - \alpha) \tag{10a}$$

$$Y = \frac{F_0}{\sqrt{[k - \omega^2(w/g)]^2 + (\omega c)^2}} (\cos \omega t \sin \beta + \sin \omega t \cos \beta)$$

$$= \frac{F_0}{\sqrt{[k - \omega^2(w/g)]^2 + (\omega c)^2}} \sin (\omega t + \beta) \tag{10b}$$

The first of these equations is the more convenient because it involves the same function (the cosine) as the excitation term in the differential equation. Hence, the phase relation between the response of the system and the disturbing force can easily be inferred. Accordingly, we shall continue with the first expression for Y.

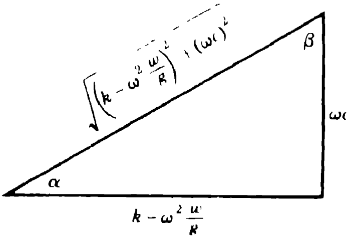

$\sqrt{\left(k - \omega^2 \frac{w}{g}\right)^2 + (\omega c)^2}$
β
ωc
α
$k - \omega^2 \frac{w}{g}$

FIGURE 5.5
The triangle defining the phase angles appearing in Eqs. (10a) and (10b).

If we divide the numerator and denominator by k and rearrange slightly, we obtain

$$Y = \frac{F_0/k}{\sqrt{[1-\omega^2(w/kg)]^2+(\omega c/k)^2}}\cos(\omega t-\alpha)$$

$$= \frac{F_0/k}{\sqrt{\left(1-\frac{\omega^2}{kg/w}\right)^2+\left(\frac{\omega}{\sqrt{kg/w}}\cdot\frac{2c}{\sqrt{4kw/g}}\right)^2}}\cos(\omega t-\alpha)$$

$$= \frac{\delta_{st}}{\sqrt{[1-(\omega^2/\omega_n^2)]^2+[2(\omega/\omega_n)(c/c_c)]^2}}\cos(\omega t-\alpha)$$

where $\delta_{st} \equiv F_0/k$ is the **static deflection** which a *constant* force of magnitude F_0 would produce in a spring of modulus k, and, as before, $\omega_n^2 = kg/w$ and $c_r^2 = 4kw/g$

The quantity

$$M = \frac{1}{\sqrt{[1-(\omega^2/\omega_n^2)]^2+[2(\omega/\omega_n)(c/c_c)]^2}} \tag{11}$$

is called the **magnification ratio.** It is the factor by which the static deflection produced in a spring of modulus k by a constant force F_0 must be multiplied in order to give the amplitude of the vibrations which result when the same force acts dynamically with frequency ω Curves of the magnification ratio M plotted against the **frequency ratio** ω/ω_n for various values of the **damping ratio** c/c_r are shown in Fig 5 6 An inspection of Fig 5 6 reveals the following interesting facts

a $M = 1$, regardless of the amount of damping, if $\omega/\omega_n = 0$

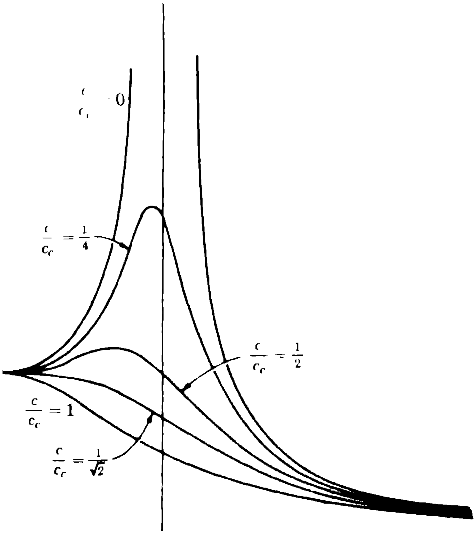


FIGURE 5 6 Curves of the magnification ratio M as a function of the impressed frequency ratio ω/ω_n for various amounts of damping

b If $0 < c/c_c < 1/\sqrt{2}$, M rises to a maximum as ω/ω_n increases from 0, the peak value of M occurring in all cases *before* the impressed frequency ω reaches the undamped natural frequency ω_n.

c The smaller the amount of friction, the larger the maximum of M, until for conditions of undamped resonance, namely, $c/c_c = 0$ and $\omega = \omega_n$, infinite magnification, i.e., a response of infinite amplitude, occurs.

d If $c/c_c \geqq 1/\sqrt{2}$, the magnification ratio decreases steadily as ω/ω_n increases from 0.

e For all values of c/c_c, M approaches zero as the impressed frequency is raised indefinitely above the undamped natural frequency of the system.

The angle
$$\alpha = \tan^{-1}\frac{\omega c}{k - \omega^2(w/g)} \qquad 0 \leqq \alpha \leqq \pi$$

which appears in Eq. (10*a*) and is shown in Fig. 5.5, is known as the **phase angle** or **angle of lag** of the response. Like the magnification ratio, it, too, can easily be expressed in terms of the dimensionless parameters ω/ω_n and c/c_c. To do this we need only divide the numerator and denominator of the right-hand side of the last expression by k and rearrange slightly

$$\alpha = \tan^{-1}\frac{\omega c/k}{1 - \omega^2(w/kg)}$$

$$= \tan^{-1}\frac{(\omega/\sqrt{kg/w})(2c/\sqrt{4kw/g})}{1 - \omega^2/(kg/w)}$$

$$= \tan^{-1}\frac{2(\omega/\omega_n)(c/c_c)}{1 - (\omega/\omega_n)^2} \tag{12}$$

It is important to note that α is *not* to be read from the principal-value branch of the arctangent function, for it is evident from Fig. 5.5 that $\sin\alpha$ is always positive, whereas $\cos\alpha$ can be either positive or negative. Hence, α must be an angle between 0 and π and not an angle in the principal-value range $(-\pi/2,\pi/2)$. Plots of α versus the frequency ratio ω/ω_n for various values of the damping ratio c/c_c are shown in Fig. 5.7.

The physical significance of α is shown in Fig. 5.8. The displacement Y reaches its maxima α/ω units of time *after* or *later than* the driving force reaches its corresponding peak values. When the frequency of the disturbing force is well below the undamped natural frequency of the system, α is small and the forced vibrations lag only slightly behind the driving force. When the impressed frequency is equal to the natural frequency, the response of the system lags the excitation by one-quarter of a cycle. As ω increases indefinitely, the lag of the response approaches half a cycle, or, in other words, the response becomes 180° out of phase with respect to the driving force.

The results of our detailed study of the vibrating weight can now be summarized. The complete motion of the system consists

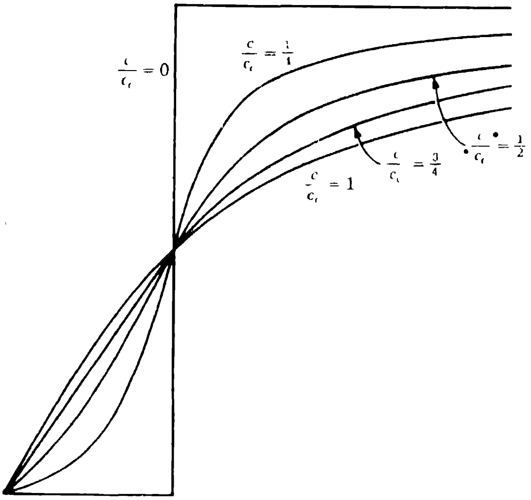


FIGURE 5.7 Curves of the phase angle α as a function of the impressed frequency ratio ω/ω_n for various amounts of damping

of two parts. The first is described by the complementary function of the underlying differential equation and may be either oscillatory or nonoscillatory according as the amount of friction in the system is less than or more than the critical damping figure for the system. In any case, however, this part of the solution contains factors which decay exponentially and thus becomes vanishingly small in a very short time. For this reason it is known as the **transient.** The general expression for the transient contains

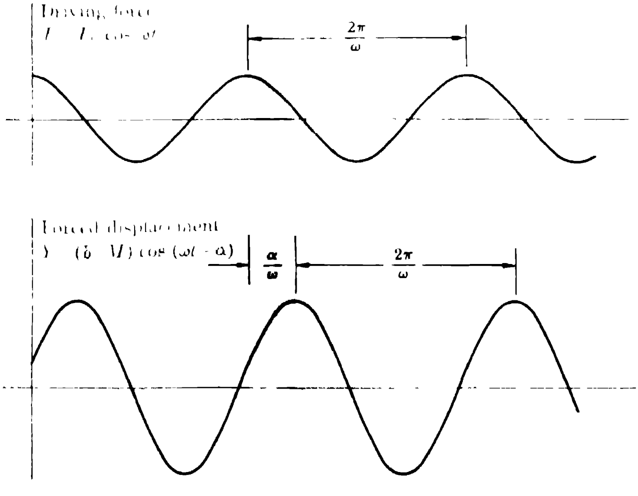


FIGURE 5.8 Plot showing the significance of the phase angle as a measure of the time by which the response lags the excitation in a mechanical system

two arbitrary constants, which, after the complete solution has been constructed, must be determined to fit the initial conditions of displacement and velocity The second part of the solution is described by the particular integral In the highly important case in which the system is acted upon by a pure harmonic disturbing force (we considered only $F = F_0 \cos \omega t$, but without exception all our conclusions are equally valid for $F = F_0 \sin \omega t$), this term represents a harmonic displacement of the same frequency as the excitation but lagging behind the latter The amplitude of this displacement is a definite multiple of the steady deflection which would be produced in the system by a constant force of the same magnitude as the actual, alternating force This factor of magnification, like the amount of lag, depends solely on the amount of friction in the system and on the ratio of the impressed frequency to the undamped natural frequency of the system The particular integral does not decay as time goes on but continues indefinitely in the same pattern For this reason it is known as the **steady state.**

EXAMPLE 1

A 50-lb weight is suspended from a spring of modulus 20 lb/in When the system is vibrating freely, it is observed that in consecutive cycles the maximum displacement decreases by 40 per cent If a force equal to 10 cos ωt acts upon the system, find the amplitude of the resultant steady-state motion if (a) $\omega = 6$, (b) $\omega = 12$, and (c) $\omega = 18$ rad/sec

The first step here is to determine the amount of damping present in the system From the given data it is clear that

$$y_{n+2} = 0.60 y_n$$

and, thus, that
$$\delta = \ln \frac{y_n}{y_{n+2}} = \ln \frac{1}{0.60} = 0.511$$

Hence, by Eq (9),

$$\frac{c}{c_c} = \frac{\delta}{\sqrt{\delta^2 + 4\pi^2}} = \frac{0.511}{\sqrt{(0.511)^2 + 4\pi^2}} = 0.081$$

Next we must compute the undamped natural frequency of the system Using Eq (6), we have

$$\omega_n = \sqrt{\frac{kg}{w}} = \sqrt{\frac{20 \times 384}{50}} = 12.4 \text{ rad/sec}$$

Knowing c/c_c and ω_n, we can now use Eq (11) to compute the magnification ratio for $\omega = 6$, 12, and 18 Direct substitution gives the values

ω	6	12	18
M	1.30	5.94	0.88

Finally, it is clear that a 10 lb force, acting statically, will stretch a spring of modulus 20 lb/in. a distance

$$\delta_{st} = {}^{10}\!/_{20} = 0.5 \text{ in}$$

Hence, multiplying this static deflection by the appropriate values of the magnification ratio, we find, for the amplitude A of the steady-state motion, the values

ω	6	12	18
A	0.65	2.97	0.44

The amplitude corresponding to the impressed frequency $\omega = 12$ is much larger than either of the others because this frequency very nearly coincides with the natural frequency of the system, $\omega_n = 12.4$.

EXAMPLE 2

A system containing a negligible amount of damping is disturbed from its equilibrium position by the sudden application at $t = 0$ of a force equal to $F_0 \sin \omega t$. Discuss the subsequent motion of the system if ω is close to the natural frequency ω_n.

The differential equation to be solved here is

$$\frac{w}{g}\frac{d^2y}{dt^2} + ky = F_0 \sin \omega t$$

The complementary function is, clearly,

$$A \cos \sqrt{\frac{kg}{w}}\, t + B \sin \sqrt{\frac{kg}{w}}\, t$$

and it is easy to verify that a particular integral is

$$Y = \frac{F_0}{k - \omega^2(w/g)} \sin \omega t$$

Hence, recalling that $\omega_n = \sqrt{kg/w}$, a complete solution can be written

$$y = A \cos \omega_n t + B \sin \omega_n t + \frac{F_0 g}{w} \frac{\sin \omega t}{\omega_n^2 - \omega^2}$$

Since $y = 0$ when $t = 0$, we must have $A = 0$, leaving

$$y = B \sin \omega_n t + \frac{F_0 g}{w} \frac{\sin \omega t}{\omega_n^2 - \omega^2} \tag{13}$$

and

$$v = \frac{dy}{dt} = B\omega_n \cos \omega_n t + \frac{F_0 g}{w} \frac{\omega \cos \omega t}{\omega_n^2 - \omega^2}$$

Substituting $v = 0$ and $t = 0$ in the last equation, we obtain

$$0 = B\omega_n + \frac{F_0 g \omega}{w(\omega_n^2 - \omega^2)} \qquad \text{or} \qquad B = -\frac{F_0 g \omega}{w(\omega_n^2 - \omega^2)\omega_n}$$

Hence, substituting into (13), we find for the required solution

$$y = \frac{F_0 g}{w(\omega_n^2 - \omega^2)} \left(-\frac{\omega}{\omega_n} \sin \omega_n t + \sin \omega t \right)$$

If the impressed frequency ω is very close to the natural frequency ω_n, we can for descriptive purposes set $\omega/\omega_n = 1$ in the last expression, obtaining

$$y = \frac{F_0 g}{w} \frac{\sin \omega_n t - \sin \omega t}{\omega^2 - \omega_n^2}$$

If we convert the difference of the sine terms into a product, we get

$$y = -\frac{F_0 g}{w}\,\frac{2\cos\left(\dfrac{\omega+\omega_n}{2}t\right)\sin\left(\dfrac{\omega-\omega_n}{2}t\right)}{(\omega+\omega_n)(\omega-\omega_n)}$$

If we now denote the small quantity $\omega - \omega_n$ by 2ϵ and note that $\omega + \omega_n$ is approximately equal to 2ω, we can write

$$y = -\frac{F_0 g}{w}\,\frac{\sin \epsilon t}{2\omega\epsilon}\cos \omega t$$

Since ϵ is a small quantity, the period $2\pi/\epsilon$ of the term $\sin \epsilon t$ is large Hence, the form of the last expression shows that y can be regarded as essentially a periodic function $\cos \omega t$ of frequency ω, with slowly varying amplitude

$$-\frac{F_0 g}{w}\,\frac{\sin \epsilon t}{2\omega\epsilon}$$

Figure 5 9 shows the general nature of this behavior when ω is nearly but not quite equal to ω_n and in the limiting case when $\omega = \omega_n$ and conditions of **pure resonance** exist

This is one of the simplest illustrations of the phenomenon of **beats,** which occurs whenever an impressed frequency is close to a natural frequency of a system or whenever two slightly different frequencies are impressed upon a system regardless of what its natural frequencies may be A wave form of variable amplitude, such as that shown in Fig 5 9*a*, is said to be **amplitude-modulated,** and the lighter curves to which the actual wave periodically rises and falls are called its **envelope.**

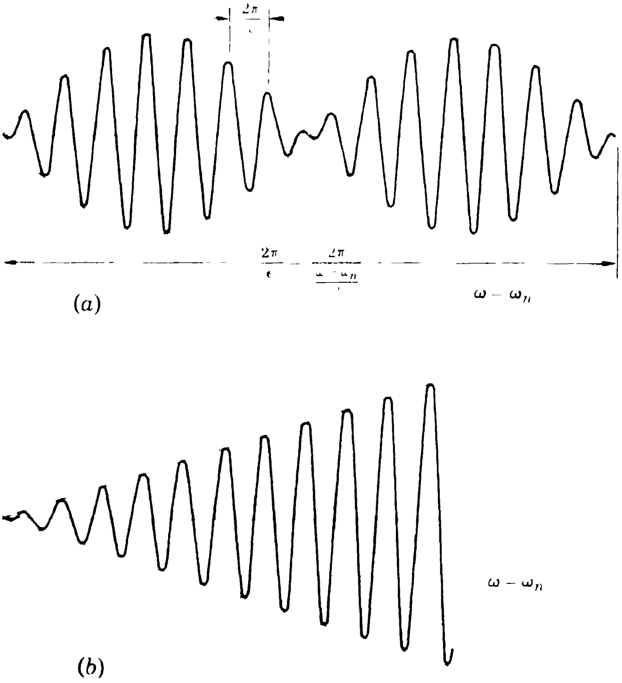


FIGURE 5 9
Plot illustrating the phenomenon of beats

EXERCISES

1. If friction is neglected, show that the natural frequency of a system consisting of a mass on an elastic suspension is approximately $3.13/\sqrt{\delta_{st}}$ cycles/sec, where δ_{st} is the deflection, in inches, produced in the suspension when the mass hangs in static equilibrium.
2. A heavy motor of unknown weight is set upon a felt mounting pad of unknown spring constant. What is the natural frequency of the system if the motor is observed to compress the pad $\frac{1}{16}$ in.?
3. Prove that the logarithmic decrement δ is equal to the natural logarithm of the ratio of any nonzero displacement to the displacement one full cycle later.
4. Show that the logarithmic decrement δ can also be computed from the formula

$$\delta = \frac{1}{k}\ln\frac{y_n}{y_{n+2k}} \qquad k = 1, 2, 3, \ldots$$

5. For a given value of c/c_c, determine the minimum number of cycles required to produce a reduction of at least 50 per cent in the amplitude of a damped oscillation.
6. If c/c_c is small, show that the logarithmic decrement is approximately

$$\delta = \frac{y_n - y_{n+2}}{y_n} = \frac{\Delta y_n}{y_n}$$

7. Show that the energy dissipated during the nth cycle of a damped oscillation is equal to $(k/2)(y_n^2 - y_{n+2}^2)$. Hence, using the result of Exercise 6, show that, when c/c_c is small, the energy loss during the nth cycle is approximately $ky_n^2\delta$.
8. If the roots of the characteristic equation in the overdamped case are $m = -r \pm s$, show that in general the complementary function can be written as $y = Ae^{-rt}\cosh(st + B)$ or as $y = Ce^{-rt}\sinh(st + D)$, according as it has no real zero or one real zero. Are there any exceptions?
9. If y_0 and v_0 are, respectively, the initial displacement and initial velocity with which an overdamped system begins its motion, show that

$$\frac{w}{g}\left(\frac{v_0}{y_0}\right)^2 + c\frac{v_0}{y_0} + k > 0$$

is the condition that the complementary function have a real zero.

10. In addition to the condition of Exercise 9, what further requirement is necessary to ensure that the zero of the complementary function will be positive, i.e., will occur during the actual motion?
11. An overdamped system begins to move from its equilibrium position with velocity v_0. Show that its maximum displacement occurs when

$$t = \frac{1}{\omega_n\sqrt{(c/c_c)^2 - 1}}\tanh^{-1}\sqrt{1 - \left(\frac{c_c}{c}\right)^2}$$

(Hint: Use the results of Exercise 8.)

12. In Exercise 11, show that the maximum displacement is

$$y_{\max} = \frac{v_0}{\omega_n}\left(\tan\frac{\alpha}{2}\right)^{\sec\alpha} \qquad \text{where } \alpha = \operatorname{Sin}^{-1}\frac{c_c}{c}$$

13. Investigate the answers to Exercises 11 and 12 in the limit when c/c_c approaches 1. Check your results by working directly with the equation for the transient in the critically damped case.

14 Show that the maximum displacements during the free motion of an underdamped system do not occur midway between the zeros of the displacement, but precede the mid-points by the constant amount

$$\frac{\operatorname{Sin}^{-1}(c/c_c)}{\omega_n\sqrt{1-(c/c_c)^2}}$$

15 Investigate the motion of a weight hanging on a spring when the disturbing force is equal to $F_0 \sin \omega t$ instead of $F_0 \cos \omega t$ In particular, show that Eqs (11) and (12) for the magnification ratio and phase angle, respectively, are still the same

16 Show that the maxima of the curves of the magnification ratio versus frequency ratio occur when

$$\frac{\omega}{\omega_n} = \sqrt{1 - 2\left(\frac{c}{c_c}\right)^2}$$

17 A weight of 128 lb hangs from a spring of modulus 75 lb/in The damping in the system is 28 per cent of critical Determine the motion of the weight if it is pulled downward 2 in from its equilibrium position and suddenly released

18 A weight of 96 lb hangs from a spring of modulus 25 lb/in The damping in the system is 60 per cent of critical Determine the motion of the weight if it is pulled downward 1 in from its equilibrium position and released with an upward velocity of 2 in /sec

19 Solve Exercise 18 if a constant force of 50 lb is suddenly applied to the system when it is at rest in its equilibrium position

20 A weight of 64 lb hangs from a spring of modulus 36 lb/in During the free motion of the system it is observed that the maximum displacement of the weight decreases to one-tenth of its value in five complete cycles of the motion Find the amplitude of the steady-state motion produced by a force equal to 6 sin 15t lb By what time interval does this steady-state motion lag the driving force?

21 A uniform bar of length l and weight w rests on two horizontal rollers whose axes are parallel and which rotate inwardly in opposition to each other with constant angular velocity Friction between the bar and each roller is assumed to be "dry," or Coulomb, that is, it is proportional to the normal force between the bar and the roller, the proportionality constant being the so-called **coefficient of friction** μ When the bar, which always remains in a line perpendicular to the axes of the rollers, is displaced slightly from a symmetrical position, it executes small oscillations in the horizontal direction Determine the period of this motion, and show how the value of μ can thus be found experimentally.

22 In many applications involving forces arising from rotating parts which have become unbalanced, the amplitude of the sinusoidal disturbing force acting on a system is not constant, but varies directly as the square of the frequency If a weight suspended from a spring is acted upon by a force of this character, determine the steady-state motion In particular, determine the form of the magnification ratio and the formula for the angle of lag

23 Show that the maxima on the plots of the magnification ratio versus the frequency ratio under the conditions of Exercise 22 always occur at values of the impressed frequency ω greater than the natural frequency of the system ω_n

24 A particle of weight w moves in a horizontal line under the influence of an elastic force equal to $-kx$, where x is the distance of the particle from the origin Friction in the system is assumed to be dry rather than viscous, that is, it is proportional to the normal force between the particle and the surface on which it moves, and does not depend on the velocity Show that the motion of the system is described by the differential equations

$$\frac{w}{g}\ddot{x} + kx = \mu w \qquad \text{when the particle is moving to the left}$$

$$\frac{w}{g}\ddot{x} + kx = -\mu w \qquad \text{when the particle is moving to the right}$$

If the body starts from rest at the point $x = x_0$, find x as a function of t. What is the decrease in amplitude per cycle? When will the body come to rest?

25 A system is acted upon by two forces

$$F_1 = A_1 \sin \omega_1 t \qquad \text{and} \qquad F_2 = A_2 \sin \omega_2 t$$

Friction, though present in the system, is so small that it can be neglected in determining the forced motion. Discuss the steady-state behavior of the system if ω_1 and ω_2 are nearly equal but if neither is close to the natural frequency of the system. In particular, show that the response consists of a term of frequency $(\omega_1 + \omega_2)/2$ whose amplitude is modulated by a term of frequency $(\omega_1 - \omega_2)/2$, and determine the limits between which the amplitude varies. Hint: After the particular integrals have been determined, note that the expression $K_1 \sin \omega_1 t + K_2 \sin \omega_2 t$ can be written

$$\frac{K_1 + K_2}{2}(\sin \omega_1 t + \sin \omega_2 t) + \frac{K_1 - K_2}{2}(\sin \omega_1 t - \sin \omega_2 t)$$

5.4

The series-electrical circuit

All the results we obtained in the last section can, after a suitable change in terminology, be applied to any of the other systems we have considered. However, the concepts central in one field are not always of equal importance in related fields, and it seems desirable to illustrate the minor differences in the application of our general theory to various classes of systems by considering one of the electrical circuits in some detail.

For the simple series circuit with an alternating impressed voltage, we derived (among several equivalent forms) the equation

$$L\frac{d^2Q}{dt^2} + R\frac{dQ}{dt} + \frac{1}{C}Q = E_0 \cos \omega t \tag{1}$$

and on comparing this with the differential equation of the vibrating weight

$$\frac{w}{g}\frac{d^2y}{dt^2} + c\frac{dy}{dt} + ky = F_0 \cos \omega t$$

we noted the correspondences

Mass $\dfrac{w}{g} \leftrightarrow$ inductance L

Friction $c \leftrightarrow$ resistance R

Spring modulus $k \leftrightarrow$ elastance $\dfrac{1}{C}$

Impressed force $F \leftrightarrow$ impressed voltage E

Displacement $y \leftrightarrow$ charge Q

Velocity $v \leftrightarrow$ current i

Extending this correspondence to the derived results by making the appropriate substitutions, we infer from the undamped natural frequency of the mechanical system

$$\omega_n = \sqrt{\frac{kg}{w}}$$

that the electrical circuit has a natural frequency

$$\Omega_n = \sqrt{\frac{1}{LC}}$$

when no resistance is present. Furthermore, the concept of critical damping

$$c_c = \sqrt{\frac{4kw}{g}}$$

leads to the concept of critical resistance

$$R_c = \sqrt{\frac{4L}{C}}$$

which determines whether the free behavior of the electrical system will be oscillatory or nonoscillatory.

The notion of magnification ratio can also be extended to the electrical case, but it is not customary to do so because the extension would relate to Q (the analogue of the displacement y), whereas in most electrical problems it is not Q but i which is the variable of interest. To see how a related concept arises in the electrical case, let us convert the particular integral Y given by Eq. (10b), Sec. 5.3, into its electrical equivalent. By direct substitution the result is found to be

$$Q = \frac{E_0 \sin(\omega t + \beta)}{\sqrt{[(1/C) - \omega^2 L]^2 + (\omega R)^2}} \qquad \beta = \text{Tan}^{-1}\frac{(1/C) - \omega^2 L}{\omega R}$$

To obtain the current i, we differentiate this, getting

$$\frac{dQ}{dt} = i = \frac{E_0\omega \cos(\omega t + \beta)}{\sqrt{[(1/C) - \omega^2 L]^2 + (\omega R)^2}}$$

or, dividing numerator and denominator by ω in the expressions for both i and β,

(2) $$i = \frac{E_0 \cos(\omega t - \delta)}{\sqrt{R^2 + [\omega L - (1/\omega C)]^2}}$$

where

(3) $$\delta = -\beta = \text{Tan}^{-1}\frac{\omega L - (1/\omega C)}{R}$$

From Eq. (2) we infer that the steady-state current produced by an alternating voltage is of the same frequency as the voltage, but differs from it in phase by

$$\frac{\delta}{\omega} \text{ units of time} \qquad \text{or} \qquad \frac{\delta/\omega}{2\pi/\omega} = \frac{\delta}{2\pi} \text{ cycles}$$

Moreover, from Eq. (3) it is clear that the numerator of tan δ (which is proportional to sin δ) can be either positive or negative, whereas the denominator of tan δ (which is proportional to cos δ) is always positive Hence δ must be an angle between $-\pi/2$ and $\pi/2$, and so the principal-value designation in Eq (3) is appropriate If δ is positive, the steady-state current *lags* the voltage; if δ is negative, the steady-state current *leads* the voltage

Furthermore, from Eq (2) we see that the amplitude of the steady-state current is obtained by dividing the amplitude of the impressed voltage E_0 by the expression

$$\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \tag{4}$$

By analogy with Ohm's law, $I = E/R$, the quantity (4) thus appears as a generalized resistance, although it is actually called the **impedance** of the circuit While not the analogue of the magnification ratio, the impedance is clearly a similar concept Since impedance is defined as

$$\frac{\text{Voltage}}{\text{Current}}$$

the mechanical quantity corresponding to this is the ratio

$$\frac{\text{Force}}{\text{Velocity}}$$

This is called the **mechanical impedance** by some writers and in certain mechanical problems has proved a useful notion

There is another approach to the problem of determining the steady-state current produced by a harmonic voltage that is well worth investigating Suppose that, given *either* $E = E_0 \cos \omega t$ *or* $E = E_0 \sin \omega t$, we write the basic differential equation (1) in the form

$$L\frac{d^2Q}{dt^2} + R\frac{dQ}{dt} + \frac{1}{C}Q = E_0e^{j\omega t} = E_0(\cos \omega t + j \sin \omega t)\dagger \tag{5}$$

This includes both possibilities for the voltage, and, if the real and the imaginary terms retain their identity throughout the analysis, then the real part of the particular integral corresponding to $E_0e^{j\omega t}$ will be the particular integral for $E_0 \cos \omega t$, and the imaginary part will be the particular integral for $E_0 \sin \omega t$

To see that this is actually the case, we must first find a particular integral of Eq (5) As usual, we do this by assuming

$$Q = Ae^{j\omega t}$$

and substituting into the differential equation This gives

$$L(-\omega^2Ae^{j\omega t}) + R(j\omega Ae^{j\omega t}) + \frac{1}{C}(Ae^{j\omega t}) = E_0e^{j\omega t}$$

---

† To avoid confusing $i = \sqrt{-1}$ with i = current, we shall throughout the rest of this chapter follow the standard practice of writing $\sqrt{-1} = j$

which will be an identity if and only if

$$A = \frac{E_0}{-\omega^2 L + j\omega R + (1/C)}$$

Hence $$Q = \frac{E_0}{j\omega R - \omega^2 L + (1/C)} e^{j\omega t}$$

From this, by differentiation, we find that

$$\frac{dQ}{dt} = i = \frac{j\omega E_0}{j\omega R - \omega^2 L + (1/C)} e^{j\omega t} = \frac{E_0}{R + j[\omega L - (1/\omega C)]} e^{j\omega t}$$

To find the real and imaginary parts of this expression, it is convenient to use the fact (Sec. 14.7) that any complex number $a + jb$ can be written in the form $a + jb = re^{j\delta}$, where the magnitude r and the angle δ of the complex number are related to the components a and b as shown in Fig. 5.10. Applied to the denominator of the second expression for i, this gives

$$R + j\left(\omega L - \frac{1}{\omega C}\right) = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}\, e^{j\delta}$$

where $$\delta = \operatorname{Tan}^{-1} \frac{\omega L - (1/\omega C)}{R}$$

Hence we can rewrite i in the form

$$\begin{aligned} i &= \frac{E_0}{\sqrt{R^2 + [\omega L - (1/\omega C)]^2}\, e^{j\delta}} e^{j\omega t} \\ &= \frac{E_0}{\sqrt{R^2 + [\omega L - (1/\omega C)]^2}} e^{j(\omega t - \delta)} \\ &= E_0 \frac{\cos(\omega t - \delta) + j \sin(\omega t - \delta)}{\sqrt{R^2 + [\omega L - (1/\omega C)]^2}} \end{aligned}$$

Comparing this with Eqs. (2) and (3), it is clear that the real part here is exactly the particular integral corresponding to $E_0 \cos \omega t$, as we derived it directly. Similarly, had we taken the trouble to work it out explicitly, we would have found for the particular integral corresponding to $E_0 \sin \omega t$ precisely the imaginary part of the last expression. Since it is much easier to find the particular integral corresponding to an exponential term than it is to find the particular integral for a cosine or sine term, the advantage of using $E_0 e^{j\omega t}$ in place of $E_0 \cos \omega t$ or $E_0 \sin \omega t$ is obvious.

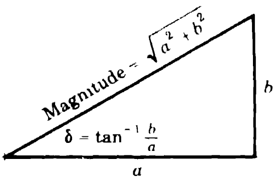

Magnitude = $\sqrt{a^2 + b^2}$
$\delta = \tan^{-1} \frac{b}{a}$
b
a

FIGURE 5.10 Plot showing the relations among the magnitude, angle, and components of a general complex number $a + jb$.

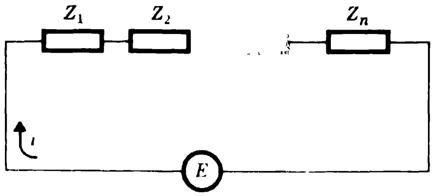


FIGURE 5.11
Impedances
nected in series

The expression

$$R + j\left(\omega L - \frac{1}{\omega C}\right) \quad \text{or} \quad j\omega L + R + \frac{1}{j\omega C},$$

is called the **complex impedance** Z. Its magnitude is the quantity (4) which we referred to simply as the impedance. Its angle δ is the **phase shift.** The real part of Z is clearly a resistance. The imaginary part of Z is called the **reactance.** The reciprocal of Z is called the **admittance.** The real part of the admittance is called the **conductance,** and the imaginary part is called the **susceptance.**

The most striking property of the complex impedance is that, when any electrical elements are connected in series or in parallel, the corresponding impedances combine just as simple resistances do. Thus the steady-state current through a series of Z's (Fig. 5.11) can be found by dividing the impressed voltage by the single impedance

$$Z = Z_1 + Z_2 + \cdots + Z_n$$

Similarly, the current through a set of elements connected in parallel (Fig. 5.12) can be found by dividing the impressed voltage by the single impedance Z, defined by the relation

$$\frac{1}{Z} = \frac{1}{Z_1} + \frac{1}{Z_2} + \cdots + \frac{1}{Z_n}$$

This makes it unnecessary to use differential equations in determining the *steady-state* behavior of an electrical network (or of a mechanical system, if the concept of mechanical impedance is used). For the *transient* behavior, however, this is not true until

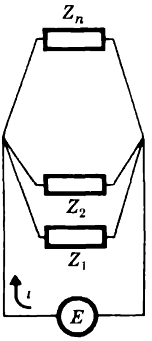


FIGURE 5 12
Impedances
connected in
parallel

the impedance concept is generalized through the use of the Laplace transformation (Chap 7).

EXAMPLE 1

A series circuit in which both the charge and the current are initially zero contains the elements $L = 1$, $R = 1{,}000$, $C = 6.25 \times 10^{-6}$. If a constant voltage $E = 24$ is suddenly switched into the circuit, find the peak value of the resultant current.

The differential equation we must solve is

$$\frac{d^2Q}{dt^2} + 1{,}000\frac{dQ}{dt} + \frac{Q}{6.25 \times 10^{-6}} = 24$$

subject to the conditions that $Q = i = 0$ when $t = 0$. The characteristic equation in this case is

$$m^2 + 1{,}000m + 160{,}000 = 0$$

and its roots are $m_1 = -200$, $m_2 = -800$. Hence, the complementary function is

$$c_1e^{-200t} + c_2e^{-800t}$$

To find a particular integral, we assume $Q = A$ and substitute into the differential equation, without difficulty getting

$$A = 150 \times 10^{-6}$$

The complete solution is, therefore,

$$Q = c_1e^{-200t} + c_2e^{-800t} + 150 \times 10^{-6}$$

and, differentiating,

$$\frac{dQ}{dt} = i = -200c_1e^{-200t} - 800c_2e^{-800t}$$

Substituting the initial conditions for Q and i gives the pair of equations

$$c_1 + c_2 + 150 \times 10^{-6} = 0 \qquad \text{and} \qquad c_1 + 4c_2 = 0$$

from which we find at once

$$c_1 = -200 \times 10^{-6} \qquad c_2 = 50 \times 10^{-6}$$

Hence, $$i = 0.04(e^{-200t} - e^{-800t})$$

To find the time when i is a maximum, we must equate to zero the time derivative of i

$$0.04(-200e^{-200t} + 800e^{-800t}) = 0$$

Dividing out $0.04 \times 800e^{-200t}$ and transposing, we have $e^{-600t} = \frac{1}{4}$, and, taking logarithms, $t = 0.0023$ sec

The maximum value of i can now be found by substituting this value of t into the general expression for i. The result is

$$i_{\max} = 0.019 \text{ amp}$$

EXERCISES

1 In Example 1, find the potential difference across each element as a function of time

2 An open series circuit contains the elements $L = 0.01$, $R = 250$, $C = 10^{-6}$. At $t = 0$, with the condenser charged to the value $Q_0 = 10^{-5}$, the circuit is closed. Find the resultant current as a function of time

3 Work Exercise 2, given that the circuit elements are $L = 6.4 \times 10^{-3}$, $R = 1.6 \times 10^{2}$, $C = 10^{-6}$.

4 Work Exercise 2, given that the circuit elements are $L = 0.01$, $R = 120$, $C = 10^{-6}$.

5 A voltage $E = 120 \cos 120\pi t$ is suddenly switched into a series circuit containing the elements $L = 1$, $R = 800$, $C = 4 \times 10^{-6}$. What is the resultant steady-state current?

6 A series circuit in which $Q_0 = i_0 = 0$ contains the elements $L = 0.15$, $R = 800$, $C = 4 \times 10^{-6}$. If a constant voltage $E = 26$ is suddenly switched into the circuit, find the resultant current as a function of time.

7 Work Exercise 6, given that the circuit elements are $L = 0.16$, $R = 800$, $C = 10^{-6}$.

8 A series circuit in which $Q_0 = i_0 = 0$ contains the elements $L = 1$, $R = 1{,}000$, $C = 4 \times 10^{-6}$. A voltage $E = 110 \sin 50\pi t$ is suddenly switched into the circuit. Find the resultant current as a function of time.

9 A series circuit in which $Q_0 = i_0 = 0$ contains the elements $L = 0.02$, $R = 250$, $C = 2 \times 10^{-6}$. A constant voltage $E = 28$ is suddenly switched into the circuit. Find the time it takes for the potential difference across the condenser to build up to one-half its final value.

10 A condenser $C = 4 \times 10^{-6}$, a resistance $R = 250$, and an inductance $L = 1$ are connected in parallel. A current source delivering a constant current $I = 0.01$ is suddenly connected across the common terminals of the elements. Find the resultant voltage as a function of time.

11 **a** Prove that, if a set of elements with impedance Z_1 is connected in series with a set of elements with impedance Z_2, then the impedance of the resultant combination is $Z_1 + Z_2$.
b Prove that, if a set of elements with impedance Z_1 is connected in parallel with a set of elements with impedance Z_2, then the impedance Z of the resultant combination is given by the formula

$$\frac{1}{Z} = \frac{1}{Z_1} + \frac{1}{Z_2}$$

12 A constant voltage is suddenly switched into a nonoscillatory RLC circuit in which $Q_0 = i_0 = 0$. Show that the potential difference across the condenser can never overshoot its final value.

13 For what value(s) of ω is the impedance $\sqrt{R^2 + [\omega L - (1/\omega C)]^2}$ a minimum? Compare this with the corresponding property of the magnification ratio. Explain.

14 If the frequency of the voltage $E_0 \cos \omega t$ impressed on a series circuit is the same as the natural frequency of the circuit, show that the amplitudes of the steady-state potential differences across the inductance and the capacitance are each equal to $E_0 R_c / 2R$.

15 Instead of using the ratio R/R_c as a dimensionless parameter in circuit analysis, it is customary to use the so-called **quality factor** Q (not to be confused with the charge Q) defined as $R_c/2R$. Express the impedance and the phase angle for a simple series circuit in terms of the resistance R, the frequency ratio Ω/Ω_n, and the quality factor Q.

5.5

Systems with several degrees of freedom

The laws of Newton and Kirchhoff, together with the theory of simultaneous linear differential equations developed in Chap. 3 and the theory of difference equations developed in Sec. 4.5, form the basis for the analysis of large classes of systems with more than one degree of freedom. The details of such applications can best be made clear through examples.

EXAMPLE 1

Assuming friction to be negligible, find the natural frequencies of the mass-spring system shown in Fig 5.13

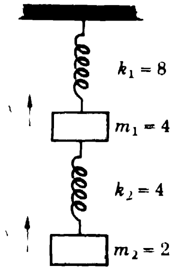


FIGURE 5.13 A simple mass-spring system

As usual, we suppose the masses to be guided, by constraints which need not be specified, so that they can move only in the vertical direction The instantaneous displacements of the masses from their equilibrium positions we shall use as coordinates to describe the system, displacements above the equilibrium positions being considered positive Since friction is assumed to be negligible, the only forces acting on the masses besides the attraction of gravity are those transmitted to them by the attached springs Moreover, as we saw in the derivation of Eq (7), Sec 5 2, the force of gravity can be neglected provided we also neglect the initial elongation of the springs and assume that each is unstretched when the system is in equilibrium

Now, when the displacements of the masses m_1 and m_2 are y_1 and y_2, respectively, the upper spring is changed in length by the amount y_1 and the lower spring is changed in length by the amount $y_1 - y_2$ Because of these changes in length, the springs exert forces equal to

$$8y_1 \qquad \text{and} \qquad 4(y_1 - y_2)$$

respectively Hence, applying Newton's law to each mass and taking due account of the direction of the forces applied to each mass by the attached springs, we have

$$4\frac{d^2y_1}{dt^2} = -8y_1 - 4(y_1 - y_2)$$

$$2\frac{d^2y_2}{dt^2} = 4(y_1 - y_2)$$

or

$$(1) \qquad \begin{aligned} (4D^2 + 12)y_1 - \qquad\quad 4y_2 &= 0 \\ -4y_1 + (2D^2 + 4)y_2 &= 0 \end{aligned}$$

From these equations we find that the equation satisfied by y_1 is

$$\begin{vmatrix} (4D^2 + 12) & -4 \\ -4 & (2D^2 + 4) \end{vmatrix} y_1 = 0$$

or

$$(D^4 + 5D^2 + 4)y_1 = 0$$

The characteristic equation of this differential equation is $m^4 + 5m^2 + 4 = 0$, and its roots are $m = \pm i, \pm 2i$. Hence,

$$(2) \qquad y_1 = c_1 \cos t + c_2 \sin t + c_3 \cos 2t + c_4 \sin 2t$$

Since the system (1) is homogeneous, it is evident that y_2 satisfies the same differential equation that y_1 satisfies Therefore,

$$(3) \qquad y_2 = d_1 \cos t + d_2 \sin t + d_3 \cos 2t + d_4 \sin 2t$$

Because we are concerned only with the frequencies at which the system can vibrate, there is no need to determine the relations which must exist between the c's and the d's. Whatever these relations may be, it is clear that Eqs. (2) and (3) represent periodic displacements at the frequencies $\omega_1 = 1$ and $\omega_2 = 2$. Moreover, since Eqs. (2) and (3) (when their coefficients are suitably related) constitute a complete solution of the system (1) and, hence, a complete description of the possible motion of the given physical system, it follows that free vibrations at frequencies other than ω_1 and ω_2 are impossible.

EXAMPLE 2

In the circuit shown in Fig. 5.14, find the current in each loop as a function of time, given that all charges and currents are zero when the switch is closed at $t = 0$.

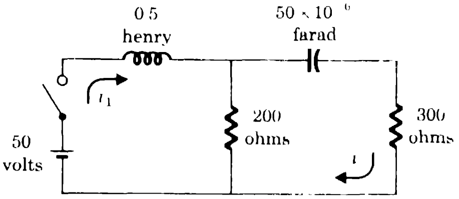


FIGURE 5.14 A simple two-loop electrical circuit.

We take as variables the currents i_1 and i_2 flowing in the respective loops, noting that the current in the common branch is, therefore, $i_1 - i_2$. Applying Kirchhoff's second law to each loop, we obtain the equations

$$0.5\frac{di_1}{dt} + 200(i_1 - i_2) = 50$$

$$300i_2 + 200(i_2 - i_1) + \frac{1}{50 \times 10^{-6}}\int i_2\,dt = 0$$

or, letting $Q_2 = \int i_2\,dt$,

$$\frac{di_1}{dt} + 400i_1 - 400\frac{dQ_2}{dt} = 100$$

$$-2i_1 + 5\frac{dQ_2}{dt} + 200Q_2 = 0$$

The characteristic equation of this system is

$$\begin{vmatrix} (m + 400) & -400m \\ -2 & (5m + 200) \end{vmatrix} = 5(m^2 + 280m + 16{,}000) = 0$$

From its roots, $m_1 = -80$ and $m_2 = -200$, we can construct the expressions

$$i_1 = a_1e^{-80t} + b_1e^{-200t}$$

$$Q_2 = a_2e^{-80t} + b_2e^{-200t}$$

which, after the constants a_1, a_2, b_1, b_2 are properly related, will constitute the complementary function of the system.

Substituting these expressions into the second of the two differential equations, we obtain

$$-2(a_1e^{-80t} + b_1e^{-200t}) + 5(-80a_2e^{-80t} - 200b_2e^{-200t}) + 200(a_2e^{-80t} + b_2e^{-200t}) = 0$$

This will be identically true if and only if $a_1 = -100a_2$ and $b_1 = -400b_2$. Therefore, the complementary function is

$$i_1 = -100a_2e^{-80t} - 400b_2e^{-200t}$$

$$Q_2 = a_2e^{-80t} + b_2e^{-200t}$$

To find a particular integral, we assume $i_1 = A_1$ and $Q_2 = A_2$. Substituting these into the nonhomogeneous system of differential equations, we obtain

$$400A_1 = 100$$

$$-2A_1 + 200A_2 = 0$$

Hence, $A_1 = \frac{1}{4}$ and $A_2 = \frac{1}{400}$

and, therefore, a complete solution of the system is

$$i_1 = -100a_2e^{-80t} - 400b_2e^{-200t} + \frac{1}{4}$$

$$Q_2 = a_2e^{-80t} + b_2e^{-200t} + \frac{1}{400}$$

Since $i_1 = 0$ and $Q_2 = 0$ when $t = 0$, we must have

$$0 = -100a_2 - 400b_2 + \frac{1}{4}$$

$$0 = a_2 + b_2 + \frac{1}{400}$$

From these we find without difficulty that $a_2 = -\frac{1}{240}$ and $b_2 = \frac{1}{600}$. The required currents are, therefore,

$$i_1 = \frac{5}{12}e^{-80t} - \frac{2}{3}e^{-200t} + \frac{1}{4}$$

$$i_2 = \frac{dQ_2}{dt} = \frac{1}{3}e^{-80t} - \frac{1}{3}e^{-200t}$$

Evidently $i_2 = 0$ when $t = 0$, as required.

EXAMPLE 3

Find the natural frequencies of the network shown in Fig. 5.15.

By applying Kirchhoff's second law to each loop in turn, we obtain the equations

$$L\frac{di_1}{dt} + \frac{1}{C}\int(i_1 - i_2)\,dt = 0$$

$$\frac{1}{C}\int(i_2 - i_1)\,dt + L\frac{di_2}{dt} + \frac{1}{C}\int(i_2 - i_3)\,dt = 0$$

$$\frac{1}{C}\int(i_{k+1} - i_k)\,dt + L\frac{di_{k+1}}{dt} + \frac{1}{C}\int(i_{k+1} - i_{k+2})\,dt = 0$$

$$\frac{1}{C}\int(i_{n-1} - i_{n-2})\,dt + L\frac{di_{n-1}}{dt} + \frac{1}{C}\int(i_{n-1} - i_n)\,dt = 0$$

$$\frac{1}{C}\int(i_n - i_{n-1})\,dt + L\frac{di_n}{dt} + \frac{1}{C}\int i_n\,dt = 0$$

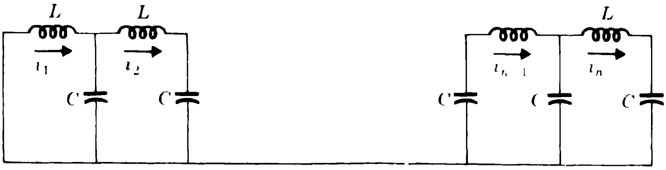


FIGURE 5.15

An oscillatory ladder-type network. (As in Fig. 4.2, although the network shown appears to contain exactly seven loops, the number of loops is actually indefinite. This is implied by the fact that the central portion of the figure is drawn with lighter lines; this convention is used throughout the book to suggest a configuration of indefinite extent.)

or, introducing new variables via the substitutions

$$\int i_k\,dt = Q_k \qquad i_k = \frac{dQ_k}{dt} = DQ_k \qquad \frac{di_k}{dt} = \frac{d^2Q_k}{dt^2} = D^2Q_k$$

and, rearranging slightly,

$$
\begin{aligned}
(LCD^2 + 1)Q_1 - Q_2 &= 0\\
-Q_1 + (LCD^2 + 2)Q_2 - Q_3 &= 0\\
&\cdot\quad\cdot\quad\cdot\\
-Q_k + (LCD^2 + 2)Q_{k+1} - Q_{k+2} &= 0 \qquad (4)\\
&\cdot\quad\cdot\quad\cdot\\
-Q_{n-2} + (LCD^2 + 2)Q_{n-1} - Q_n &= 0\\
-Q_{n-1} + (LCD^2 + 2)Q_n &= 0
\end{aligned}
$$

Since there is no resistance anywhere in the network, it is evident that the response of the circuit to any set of nonzero initial conditions of charge and current will be purely oscillatory Hence, we assume solutions of the form

$$Q_k = a_k \cos \omega t$$

where ω is the unknown frequency of the response and the a's are arbitrary constants Substituting into the equations of the set (4), dividing each equation by $-\cos \omega t$, and setting

$$LC\omega^2 = \alpha^2$$

we obtain the algebraic equations

$$
\begin{aligned}
-(1 - \alpha^2)a_1 + a_2 &= 0\\
a_1 - (2 - \alpha^2)a_2 + a_3 &= 0\\
&\cdot\quad\cdot\quad\cdot\\
a_k - (2 - \alpha^2)a_{k+1} + a_{k+2} &= 0 \qquad (5)\\
&\cdot\quad\cdot\quad\cdot\\
a_{n-2} - (2 - \alpha^2)a_{n-1} + a_n &= 0\\
a_{n-1} - (2 - \alpha^2)a_n &= 0
\end{aligned}
$$

In order for these equations to have a nontrivial solution, it is necessary that the determinant of their coefficients be zero However, in this case the determinant of the coefficients is of the nth order, and to expand it and then solve the resulting nth-degree equation in $\alpha^2 = LC\omega^2$ would be prohibitively time-consuming Hence it is much better to proceed in the following way With the exception of the first and last equations, each equation of the system (5) is of the form

$$a_k - (2 - \alpha^2)a_{k+1} + a_{k+2} = 0$$

In other words, for $k = 1, 2, \ldots, n - 2$, the a's satisfy the linear, constant-coefficient, second-order difference equation*

$$[E^2 - (2 - \alpha^2)E + 1]a_k = 0 \qquad (6)$$

The first and last equations, which clearly do not fit into the pattern of Eq (6), are, of course, the two boundary conditions necessary for the determination of the arbitrary constants which appear in the complete solution of this difference equation

---

* This is true, of course, only because the loops of the network, with the exception of the first and the last, are all identical In general, the possibility of using difference equations should always be considered in studying systems, both electrical and mechanical, which consist essentially of a number of identical components, identically connected

Following the theory of Sec 4 5, the first step in the solution of Eq (6) is to solve its characteristic equation

$$m^2 - (2 - \alpha^2)m + 1 = 0 \tag{7}$$

getting

$$m_1, m_2 = 1 - \frac{\alpha^2}{2} \pm \sqrt{\left(1 - \frac{\alpha^2}{2}\right)^2 - 1}$$

The continuation now involves an investigation of various special cases depending on the possible values of $1 - (\alpha^2/2)$

First of all, we can immediately reject the possibility that $1 - (\alpha^2/2) \geq 1$, for this implies that $\alpha^2 \leq 0$, which is impossible, since $\alpha^2 = LC\omega^2$ is an intrinsically positive quantity Moreover, if $1 - (\alpha^2/2) = -1$, that is, if $\alpha^2 = 4$, then $m_1 = m_2 = -1$, and so, according to Table 4 2, Sec 4 5, the complete solution of Eq (6) is

$$a_k = (c_1 + c_2k)(-1)^k$$

Imposing the boundary conditions on a_k, by substituting a_k into the first and last of the equations (5), we have

$$-(-3)[(c_1 + c_2)(-1)] + [(c_1 + 2c_2)(-1)^2] = -2c_1 - c_2 = 0$$

$$\{[c_1 + (n - 1)c_2](-1)^{n-1}\} - (-2)[(c_1 + nc_2)(-1)^n] = (-1)^n[c_1 + (n + 1)c_2] = 0$$

But these two equations obviously have only the trivial solution $c_1 = c_2 = 0$ Hence, $1 - (\alpha^2/2)$ cannot equal -1

If $1 - (\alpha^2/2) < -1$, we can write

$$1 - \frac{\alpha^2}{2} = -\cosh\mu \qquad \mu \neq 0 \tag{8}$$

so that the roots of the characteristic equation (7) become

$$-\cosh\mu \pm \sqrt{\cosh^2\mu - 1} = -\cosh\mu + \sinh\mu = -e^{\pm\mu}$$

Hence, a complete solution of (6) can be written

$$a_k = c_1(-e^{\mu})^k + c_2(-e^{-\mu})^k = (-1)^k(d_1\cosh\mu k + d_2\sinh\mu k)$$

where $d_1 = c_1 + c_2$ and $d_2 = c_1 - c_2$ Again imposing the boundary conditions on a_k, we have

$$-(1 + 2\cosh\mu)(d_1\cosh\mu + d_2\sinh\mu) + (d_1\cosh 2\mu + d_2\sinh 2\mu) = 0$$

$$(-1)^{n-1}[d_1\cosh(n - 1)\mu + d_2\sinh(n - 1)\mu] + (-1)^n 2\cosh\mu(d_1\cosh n\mu + d_2\sinh n\mu) = 0$$

From these, by collecting terms and then simplifying through the use of the identities

$$2\cosh^2\mu = \cosh 2\mu + 1$$

$$2\sinh\mu\cosh\mu = \sinh 2\mu$$

$$2\cosh n\mu\cosh\mu = \cosh(n + 1)\mu + \cosh(n - 1)\mu$$

$$2\sinh n\mu\cosh\mu = \sinh(n + 1)\mu + \sinh(n - 1)\mu$$

we obtain

$$(1 + \cosh\mu)d_1 + (\sinh\mu)d_2 = 0$$

$$[\cosh(n + 1)\mu]d_1 + [\sinh(n + 1)\mu]d_2 = 0$$

These equations will have a nontrivial solution if and only if

$$\begin{vmatrix} 1 + \cosh\mu & \sinh\mu \\ \cosh(n + 1)\mu & \sinh(n + 1)\mu \end{vmatrix} = \sinh(n + 1)\mu + \sinh n\mu$$

$$= 2\sinh\frac{2n + 1}{2}\mu\cosh\frac{\mu}{2} = 0$$

This can vanish only if $\mu = 0$, which is impossible, since, from (8), $\mu = 0$ implies $1 - (\alpha^2/2) = -1$, and this possibility has already been considered and rejected Hence the assumption $1 - (\alpha^2/2) < -1$ also leads only to a trivial solution

It remains now to consider the possibility $-1 < 1 - (\alpha^2/2) < 1$ To investigate this case, let us put

$$1 - \frac{\alpha^2}{2} = \cos \mu \qquad 0 < \mu < \pi \tag{9}$$

Then the roots of the characteristic equation (7) are

$$\cos \mu \pm \sqrt{\cos^2 \mu - 1} = \cos \mu \pm i \sin \mu = e^{\pm i\mu}$$

and a complete solution for a_k is

$$a_k = c_1 \cos k\mu + c_2 \sin k\mu$$

Again imposing the boundary conditions on a_k, we have

$$-(2 \cos \mu - 1)(c_1 \cos \mu + c_2 \sin \mu) + (c_1 \cos 2\mu + c_2 \sin 2\mu) = 0$$

$$[c_1 \cos (n - 1)\mu + c_2 \sin (n - 1)\mu] - 2 \cos \mu(c_1 \cos n\mu + c_2 \sin n\mu) = 0$$

From these, by collecting terms and then simplifying through the use of the identities

$$2 \cos^2 \mu = 1 + \cos 2\mu$$

$$2 \sin \mu \cos \mu = \sin 2\mu$$

$$2 \cos n\mu \cos \mu = \cos (n + 1)\mu + \cos (n - 1)\mu$$

$$2 \sin n\mu \cos \mu = \sin (n + 1)\mu + \sin (n - 1)\mu$$

we obtain

$$(\cos \mu - 1)c_1 + (\sin \mu)c_2 = 0$$

$$[\cos (n + 1)\mu]c_1 + [\sin (n + 1)\mu]c_2 = 0$$

These two equations will have a nontrivial solution for c_1 and c_2 if and only if

$$\begin{vmatrix} \cos \mu - 1 & \sin \mu \\ \cos (n + 1)\mu & \sin (n + 1)\mu \end{vmatrix} = \sin n\mu - \sin (n + 1)\mu = -2 \sin \frac{\mu}{2} \cos \frac{2n + 1}{2} \mu = 0$$

Now $\sin \mu/2$ can be zero only if μ is a multiple of 2π, which is impossible in the present case, since $0 < \mu < \pi$ Hence, we must have

$$\cos \frac{2n + 1}{2} \mu = 0$$

Therefore,

$$\frac{2n + 1}{2} \mu = \frac{2N + 1}{2} \pi$$

and

$$\mu = \frac{2N + 1}{2n + 1} \pi \qquad N = 0, 1, 2, \ldots, n - 1$$

The values $N = 0, 1, \ldots, n - 1$ lead to distinct values of μ which in turn define the n natural frequencies of the network, since, from (9),

$$\sqrt{LC\omega^2} \equiv \alpha = \sqrt{2(1 - \cos \mu)} = 2 \sin \frac{\mu}{2}$$

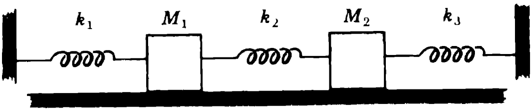
Hence the required frequencies are given by the formula

$$\omega_N = \frac{2}{\sqrt{LC}} \sin \left(\frac{2N + 1}{2n + 1} \frac{\pi}{2} \right) \qquad N = 0, 1, \ldots, n - 1$$

EXERCISES

1 If $M_1 = 1, M_2 = 2, k_1 = 1, k_2 = k_3 = 2$ for the system shown in Fig. 5.16, find the natural frequencies of the system.

FIGURE 5.16



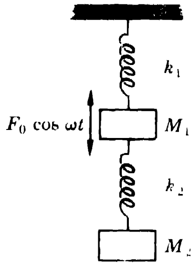
2 If $M_1 = 1, M_2 = 3, k_1 = 1, k_2 = k_3 = 3$ for the system shown in Fig. 5.16, find the natural frequencies of the system.

3 If $M_1 = M_2 = 1, k_1 = 1, k_2 = 3, k_3 = 9$ for the system shown in Fig. 5.16, find the natural frequencies of the system.

4 Find the displacements of M_1 and M_2 as functions of t in Exercise 1, if the system starts from rest with $x_1 = 1$ and $x_2 = 0$.

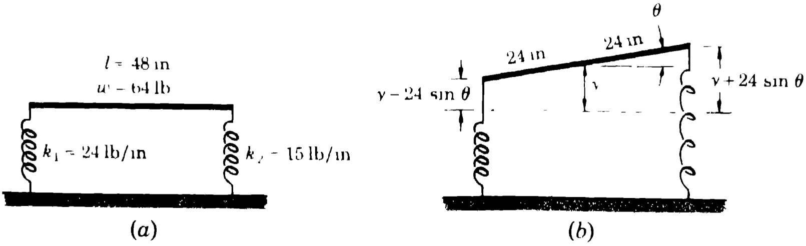
5 In the system shown in Fig. 5.17 the parameters M_1, k_1, and ω are assumed to be known. Determine k_2 and M_2 so that in the steady-state forced motion of the system the mass M_1 will remain at rest.

FIGURE 5.17



6 Prove that for no values of the parameters M_1, M_2, k_1, k_2, k_3 can the two natural frequencies of the system shown in Fig. 5.16 be equal.

7 A uniform bar 4 ft long and weighing 16 lb/ft is supported as shown in Fig. 5.18, on springs of moduli 24 and 15 lb/in., respectively. If the springs are guided so that only vertical displacement of the center of the bar is possible, find the natural frequencies of the system.



(a) (b)

FIGURE 5.18

(Hint: As coordinates, use the displacement y of the center of the bar and the angle of rotation θ of the bar about its center. Assume displacements so small that $\cos \theta$ can be replaced by 1 and $\sin \theta$ can be replaced by θ.)

8 In the network shown in Fig 5 19 the current and the charge on the condenser in the closed loop are both zero, but the condenser in the open loop bears a charge Q_0 Find the current in each loop as a function of time after the switch is closed

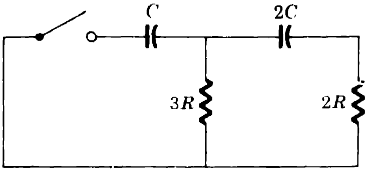


FIGURE 5 19

9 Work Exercise 8 for the network shown in Fig 5 20

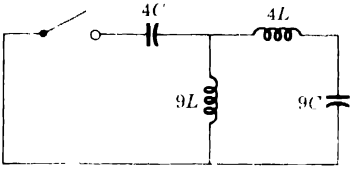


FIGURE 5 20

10 Find the current in each loop of the network shown in Fig 5 21 if the switch is closed at an instant when all charges and currents are zero

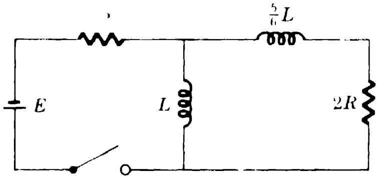


FIGURE 5 21

11 Work Exercise 10 for the network shown in Fig 5 22

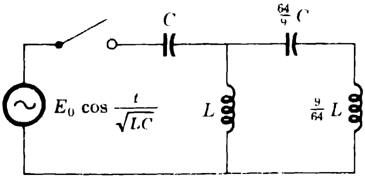


FIGURE 5 22

12 In Example 3, find the normal modes, i e , the sets of a's for each of the natural frequencies

13 Work Example 3, with the condenser in series with the inductance in the last loop removed

14 Work Example 3, with the inductances and capacitances in each loop interchanged

15 Find the natural frequencies of the system of n equal masses connected by identical springs shown in Fig 5 23

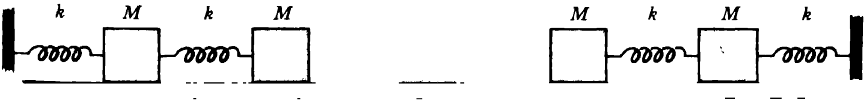


FIGURE 5 23

16 Find the natural frequencies of the system of n identical disks connected by identical lengths of elastic shafting shown in Fig 5 24

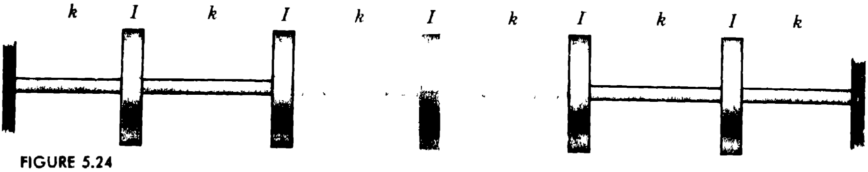


FIGURE 5.24

17 Work Exercise 15, with the spring connecting the right-hand mass to the wall removed.

18 Work Exercise 16, with the shaft connecting the right-hand disk to the wall removed

19 If a voltage $E_0 \cos \omega t$ is inserted in series with the inductance in the first loop of the network in Example 3, find expressions for the steady-state charges on the various condensers if

a $0 < \omega < \dfrac{2}{\sqrt{LC}}$ **b** $\omega > \dfrac{2}{\sqrt{LC}}$

20 If the capacitances and inductances in the network in Example 3 are interchanged and if a voltage $E_0 \cos \omega t$ is then inserted in series with the capacitance in the first loop, find expressions for the steady-state charges on the various condensers if

a $0 < \omega < \dfrac{1}{2\sqrt{LC}}$ **b** $\omega > \dfrac{1}{2\sqrt{LC}}$

CHAPTER SIX

Fourier Series and Integrals

6.1

Introduction

In Chap 2 we learned that nonhomogeneous, linear, constant-coefficient differential equations containing terms of the form

$$A \cos \omega t \qquad \text{and} \qquad B \sin \omega t$$

could easily be solved for all values of ω Then in Chap 5 we discovered that such differential equations were fundamental in the study of physical systems subjected to periodic disturbances In many cases, however, the forces, torques, voltages, or currents which act on a system, although periodic, are by no means so simple as pure sine and cosine waves For instance, the voltage impressed on an electrical circuit might consist of a series of pulses as shown in Fig 6 1*a*, or the disturbing influence acting on a mechanical system might be a force of constant magnitude whose direction is periodically and instantaneously reversed, as in Fig 6 1*b*

This raises the question of whether or not a general periodic

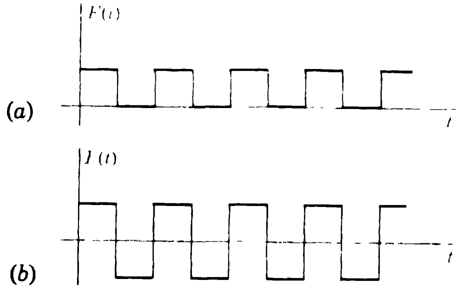


FIGURE 6.1 Typical periodic forcing functions

function* can be expressed as a series of sine and cosine terms. Specifically, since, for all integral values of n,

$$\cos\frac{n\pi(t+2p)}{p} = \cos\frac{n\pi t}{p} \qquad \text{and} \qquad \sin\frac{n\pi(t+2p)}{p} = \sin\frac{n\pi t}{p}$$

it is natural to ask whether an arbitrary function $f(t)$ of period $2p$ can be represented by a series of the form

$$\frac{a_0}{2}^{\dagger} + a_1\cos\frac{\pi t}{p} + a_2\cos\frac{2\pi t}{p} + \cdots + a_n\cos\frac{n\pi t}{p} + \cdots$$
$$+ b_1\sin\frac{\pi t}{p} + b_2\sin\frac{2\pi t}{p} + \cdots + b_n\sin\frac{n\pi t}{p} + \cdots$$

If this is the case, then the methods of Chap. 5, applied to the individual terms of such a series, will enable us, in fact, to analyze the behavior of systems acted upon by general periodic disturbances. The possibility of such expansions and their determination when they exist are the subject matter of **Fourier analysis,**‡ to which we shall devote this chapter.

6.2

The Euler coefficients

To obtain formulas for the coefficients a_n and b_n in the expansion

(1) $$f(t) = \frac{a_0}{2} + a_1\cos\frac{\pi t}{p} + a_2\cos\frac{2\pi t}{p} + \cdots + a_n\cos\frac{n\pi t}{p} + \cdots$$
$$+ b_1\sin\frac{\pi t}{p} + b_2\sin\frac{2\pi t}{p} + \cdots + b_n\sin\frac{n\pi t}{p} + \cdots$$

assuming, of course, that it exists, we shall need the following definite integrals, which are valid for all values of d, provided m and n are integers satisfying the given restrictions.

(2) $$\int_d^{d+2p} \cos\frac{n\pi t}{p}\,dt = 0 \qquad n \neq 0$$

(3) $$\int_d^{d+2p} \sin\frac{n\pi t}{p}\,dt = 0$$

---

* A function $f(t)$ is said to be **periodic** if there exists a constant $2p$ with the property that

$f(t + 2p) = f(t)$ for all t

If $2p$ is the smallest number for which this identity holds, it is called the **period** of the function.

† The introduction of the factor $\frac{1}{2}$ is a conventional device to render more symmetrical the final formulas for the coefficients.

‡ Named for Joseph Fourier (1768–1830), French mathematician and confidant of Napoleon, who first undertook the systematic study of such expansions in a memorable monograph, "Théorie analytique de la chaleur," published in 1822. The use of such series in particular problems, however, dates from the time of Daniel Bernoulli (1700–1782), who used them to solve certain problems connected with vibrating strings.

$$\int_d^{d+2p} \cos\frac{m\pi t}{p}\cos\frac{n\pi t}{p}\,dt = 0 \qquad m \neq n \tag{4}$$

$$\int_d^{d+2p} \cos^2\frac{n\pi t}{p}\,dt = p \qquad n \neq 0 \tag{5}$$

$$\int_d^{d+2p} \cos\frac{m\pi t}{p}\sin\frac{n\pi t}{p}\,dt = 0 \tag{6}$$

$$\int_d^{d+2p} \sin\frac{m\pi t}{p}\sin\frac{n\pi t}{p}\,dt = 0 \qquad m \neq n \tag{7}$$

$$\int_d^{d+2p} \sin^2\frac{n\pi t}{p}\,dt = p \qquad n \neq 0 \tag{8}$$

With these integrals available, the determination of a_n and b_n proceeds as follows.*

To find a_0, we assume that the series (1) can legitimately be integrated term by term from $t = d$ to $t = d + 2p$.† Then

$$\int_d^{d+2p} f(t)\,dt = \frac{a_0}{2}\int_d^{d+2p} dt + a_1\int_d^{d+2p}\cos\frac{\pi t}{p}\,dt + \cdots$$
$$+ a_n\int_d^{d+2p}\cos\frac{n\pi t}{p}\,dt + \cdots$$
$$+ b_1\int_d^{d+2p}\sin\frac{\pi t}{p}\,dt + \cdots$$
$$+ b_n\int_d^{d+2p}\sin\frac{n\pi t}{p}\,dt + \cdots$$

The integral on the left can always be evaluated, since $f(t)$ is a known function which is assumed to be integrable. At worst, some method of approximate integration, such as those we discussed in Sec. 4.3, will be required. The first term on the right is simply

$$\frac{1}{2}a_0 t\Big|_d^{d+2p} = pa_0$$

By Eq. (2), all integrals with a cosine in the integrand vanish, and, by Eq. (3), all integrals containing a sine vanish. Hence, the integrated result reduces to

$$\int_d^{d+2p} f(t)\,dt = pa_0$$

or

$$a_0 = \frac{1}{p}\int_d^{d+2p} f(t)\,dt \tag{9}$$

To find a_n ($n = 1, 2, 3, \ldots$), we multiply each side of (1) by $\cos n\pi t/p$ and then integrate from d to $d + 2p$, assuming again that term-by-term integration is justified. This gives

---

* The procedure here is analogous to the procedure we used in Sec. 4.6 to express an arbitrary polynomial as a linear combination of orthogonal polynomials. The most obvious difference is that in Sec. 4.6 the orthogonality condition involved the *summation* of products of two different functions over a discrete set of values, but here the orthogonality condition involves the *integration* of products of two different functions.

† A sufficient condition for a series of integrable functions to be term-by-term integrable is that it be uniformly convergent (Theorem 6, Sec. 15.1).

$$\begin{aligned}\int_d^{d+2p} f(t)\cos\frac{n\pi t}{p}\,dt = \frac{1}{2}a_0\int_d^{d+2p}\cos\frac{n\pi t}{p}\,dt \\ + a_1\int_d^{d+2p}\cos\frac{\pi t}{p}\cos\frac{n\pi t}{p}\,dt + \cdots \\ + a_n\int_d^{d+2p}\cos^2\frac{n\pi t}{p}\,dt + \cdots \\ + b_1\int_d^{d+2p}\sin\frac{\pi t}{p}\cos\frac{n\pi t}{p}\,dt + \cdots \\ + b_n\int_d^{d+2p}\sin\frac{n\pi t}{p}\cos\frac{n\pi t}{p}\,dt + \cdots\end{aligned}$$

Again the integral on the left is completely determined By Eqs. (2) and (4), all integrals on the right containing only cosine terms vanish except the one involving $\cos^2 n\pi t/p$, which, by Eq (5), is equal to p Finally, by Eq (6), every integral which contains a sine is zero. Hence,

$$\int_d^{d+2p} f(t)\cos\frac{n\pi t}{p}\,dt = pa_n$$

or

$$a_n = \frac{1}{p}\int_d^{d+2p} f(t)\cos\frac{n\pi t}{p}\,dt \tag{10}$$

To determine b_n, we continue essentially the same procedure We multiply (1) by $\sin n\pi t/p$ and then integrate from d to $d + 2p$, getting

$$\begin{aligned}\int_d^{d+2p} f(t)\sin\frac{n\pi t}{p}\,dt = \frac{1}{2}a_0\int_d^{d+2p}\sin\frac{n\pi t}{p}\,dt \\ + a_1\int_d^{d+2p}\cos\frac{\pi t}{p}\sin\frac{n\pi t}{p}\,dt + \cdots \\ + a_n\int_d^{d+2p}\cos\frac{n\pi t}{p}\sin\frac{n\pi t}{p}\,dt + \cdots \\ + b_1\int_d^{d+2p}\sin\frac{\pi t}{p}\sin\frac{n\pi t}{p}\,dt + \cdots \\ + b_n\int_d^{d+2p}\sin^2\frac{n\pi t}{p}\,dt + \cdots\end{aligned}$$

As before, every integral on the right vanishes but one, leaving

$$\int_d^{d+2p} f(t)\sin^2\frac{n\pi t}{2}\,dt = pb_n$$

or

$$b_n = \frac{1}{p}\int_d^{d+2p} f(t)\sin\frac{n\pi t}{p}\,dt \tag{11}$$

Formulas (9), (10), and (11) are known as the **Euler** or **Euler-Fourier formulas,** and the series (1), when its coefficients have these values, is known as the **Fourier series** of $f(t)$ In most applications, the interval over which the coefficients are computed is either $(-p,p)$ or $(0,2p)$, so the value of d in the Euler formulas is usually either $-p$ or 0 Actually, the formula for a_0 need not be listed, for it can be obtained from the general expression for a_n by putting $n = 0$ † It was to achieve this that we wrote the constant term as $\frac{1}{2}a_0$ in the original expansion

We must be careful at this stage not to delude ourselves with the belief that we have proved that every periodic function $f(t)$ has a Fourier expansion which converges to it What our analysis has shown is merely that *if* a function $f(t)$ has an expansion of the form (1) for which term-by-term integration is valid, *then* the coefficients in that series must be given by the Euler formulas Questions concerning the convergence of Fourier series and, if they converge, the conditions under which they will represent the functions which generated them are many and difficult and are by no means completely answered yet These problems are primarily of theoretical interest, however, for almost any conceivable practical application is covered by the famous **theorem of Dirichlet**:‡

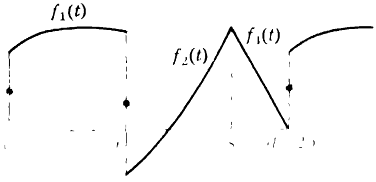
THEOREM 1

If $f(t)$ is a bounded periodic function which in any one period has at most a finite number of local maxima and minima and a finite number of points of discontinuity, then the Fourier series of $f(t)$ converges to $f(t)$ at all points where $f(t)$ is continuous and converges to the average of the right- and left-hand limits of $f(t)$ at each point where $f(t)$ is discontinuous

The conditions of Theorem 1, which are usually referred to as the **Dirichlet conditions**, make it clear that a function need not be continuous in order to possess a valid Fourier expansion This means that a function may have a graph consisting of a number of disjointed arcs of different curves, each defined by a different formula, and still be representable by a Fourier series In using the Euler formulas to find the coefficients in the expansion of such a function it will, therefore, be necessary to break up the range of integration $(d, d + 2p)$ to correspond to the various segments of the function Thus, in Fig 6 2, the function $f(t)$ is

FIGURE 6 2

A periodic function defined by different formulas over different portions of a period



---

† It is not necessarily the case, however, that the value of a_0 in a particular problem can be obtained by putting $n = 0$ in the *integrated* formula for a_n For instance, in Example 2, the integrated formula for a_n is indeterminate when $n = 0$, and evaluation of the indeterminacy yields -3, instead of the correct value 3 which is obtained by putting $n = 0$ *before* integrating

‡ Named for the German mathematician Peter Gustave Lejeune Dirichlet (1805–1859) For a proof of this theorem see, for instance, H S Carslaw, "Fourier Series," pp. 225–232, Dover Publications, Inc , New York, 1930.

defined by three different expressions $f_1(t)$, $f_2(t)$, $f_3(t)$, on successive portions of the period interval $d \leqq t \leqq d + 2p$. Hence it is necessary to write the Euler formulas as

$$\begin{aligned}
a_n &= \frac{1}{p}\int_d^{d+2p} f(t)\cos\frac{n\pi t}{p}\,dt \\
&= \frac{1}{p}\int_d^r f_1(t)\cos\frac{n\pi t}{p}\,dt + \frac{1}{p}\int_r^s f_2(t)\cos\frac{n\pi t}{p}\,dt + \frac{1}{p}\int_s^{d+2p} f_3(t)\cos\frac{n\pi t}{p}\,dt \\
b_n &= \frac{1}{p}\int_d^{d+2p} f(t)\sin\frac{n\pi t}{p}\,dt \\
&= \frac{1}{p}\int_d^r f_1(t)\sin\frac{n\pi t}{p}\,dt + \frac{1}{p}\int_r^s f_2(t)\sin\frac{n\pi t}{p}\,dt + \frac{1}{p}\int_s^{d+2p} f_3(t)\sin\frac{n\pi t}{p}\,dt
\end{aligned}$$

Incidentally, according to Theorem 1, the Fourier series of the function shown in Fig. 6.2 will converge to the average values, indicated by dots, at the discontinuities at d, r, and $d + 2p$, regardless of the definition (or lack of definition) of the function at these points.

EXAMPLE 1

What is the Fourier expansion of the periodic function whose definition in one period is

$$f(t) = \begin{cases} 0 & -\pi < t < 0 \\ \sin t & 0 < t < \pi \end{cases}$$

In this case the half period of the given function is $p = \pi$. Hence, taking $d = -\pi$ in the Euler formulas, we have

$$\begin{aligned}
a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt\,dt = \frac{1}{\pi}\int_{-\pi}^{0} 0\,\cos nt\,dt + \frac{1}{\pi}\int_0^{\pi}\sin t\cos nt\,dt \\
&= \frac{1}{\pi}\left[-\frac{1}{2}\left(\frac{\cos(1-n)t}{1-n} + \frac{\cos(1+n)t}{1+n}\right)\right]_0^{\pi} \\
&= -\frac{1}{2\pi}\left[\frac{\cos(\pi-n\pi)}{1-n} + \frac{\cos(\pi+n\pi)}{1+n} - \left(\frac{1}{1-n} + \frac{1}{1+n}\right)\right] \\
&= -\frac{1}{2\pi}\left(\frac{-\cos n\pi}{1-n} + \frac{-\cos n\pi}{1+n} - \frac{2}{1-n^2}\right) \\
&= \frac{\cos n\pi + 1}{\pi(1-n^2)} \qquad n \neq 1 \\
a_1 &= \frac{1}{\pi}\int_0^{\pi}\sin t\cos t\,dt = \frac{\sin^2 t}{2\pi}\bigg|_0^{\pi} = 0 \\
b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt\,dt = \frac{1}{\pi}\int_{-\pi}^{0} 0\,\sin nt\,dt + \frac{1}{\pi}\int_0^{\pi}\sin t\sin nt\,dt \\
&= \frac{1}{\pi}\left[\frac{1}{2}\left(\frac{\sin(1-n)t}{1-n} - \frac{\sin(1+n)t}{1+n}\right)\right]_0^{\pi} = 0 \qquad n \neq 1 \\
b_1 &= \frac{1}{\pi}\int_0^{\pi}\sin^2 t\,dt = \frac{1}{\pi}\left[\frac{t}{2} - \frac{\sin 2t}{4}\right]_0^{\pi} = \frac{1}{2}
\end{aligned}$$

Hence, evaluating the coefficients for $n = 0, 1, 2, \ldots$, we have

$$f(t) = \frac{1}{\pi} + \frac{\sin t}{2} - \frac{2}{\pi}\left(\frac{\cos 2t}{3} + \frac{\cos 4t}{15} + \frac{\cos 6t}{35} + \frac{\cos 8t}{63} + \cdots\right)$$

Plots showing the accuracy with which the first n terms of this series represent the given function are shown in Fig. 6.3 for $n = 1, 2, 3$. For $n = 4, 5, \ldots$, the graphs of the partial sums are almost indistinguishable from the graph of $f(t)$.

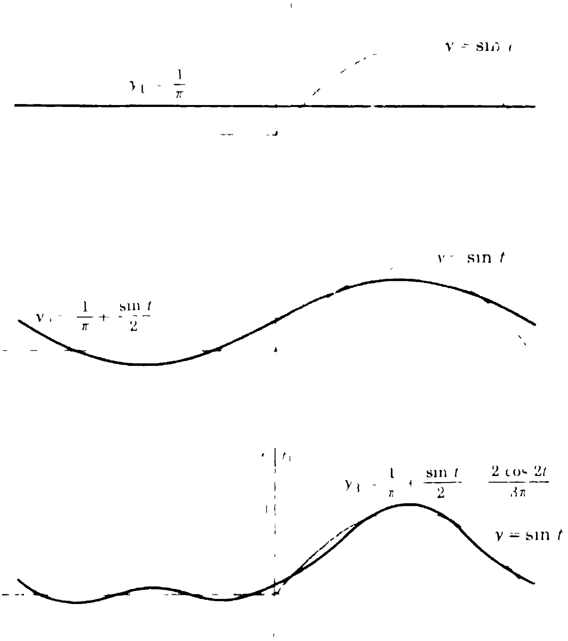


FIGURE 6.3 The approximation of a function by the first few terms of its Fourier expansion.

Interesting numerical series can often be obtained from Fourier series by evaluating them at specific points. For instance, if we set $t = \pi/2$ in the above expansion, we find

$$f\left(\frac{\pi}{2}\right) = 1 = \frac{1}{\pi} + \frac{1}{2} - \frac{2}{\pi}\left(-\frac{1}{3} + \frac{1}{15} - \frac{1}{35} + \frac{1}{63} - \cdots\right)$$

or

$$\frac{1}{1\cdot 3} - \frac{1}{3\cdot 5} + \frac{1}{5\cdot 7} - \frac{1}{7\cdot 9} + \cdots = \frac{\pi - 2}{4}$$

EXAMPLE 2

Find the Fourier expansion of the periodic function whose definition in one period is

$$f(t) = \begin{cases} -t & -3 < t < 0 \\ t & 0 < t < 3 \end{cases}$$

In this case the period of the function is 6. Hence $p = 3$, and, from (10) and (11), taking

$d = -3$, we have

$$
\begin{aligned}
a_n &= \frac{1}{3}\int_{-3}^{0} -t\cos\frac{n\pi t}{3}\,dt + \frac{1}{3}\int_{0}^{3} t\cos\frac{n\pi t}{3}\,dt \\
&= -\frac{1}{3}\left[\frac{9}{n^2\pi^2}\cos\frac{n\pi t}{3} + \frac{3t}{n\pi}\sin\frac{n\pi t}{3}\right]_{-3}^{0} + \frac{1}{3}\left[\frac{9}{n^2\pi^2}\cos\frac{n\pi t}{3} + \frac{3t}{n\pi}\sin\frac{n\pi t}{3}\right]_{0}^{3} \\
&= \frac{-3}{n^2\pi^2}(1 - \cos n\pi) + \frac{3}{n^2\pi^2}(\cos n\pi - 1) \\
&= \frac{6}{n^2\pi^2}(\cos n\pi - 1) \qquad n \neq 0
\end{aligned}
$$

$$
a_0 = \frac{1}{3}\int_{-3}^{0} -t\,dt + \frac{1}{3}\int_{0}^{3} t\,dt = -\frac{t^2}{6}\Big|_{-3}^{0} + \frac{t^2}{6}\Big|_{0}^{3} = \frac{3}{2} + \frac{3}{2} = 3
$$

$$
\begin{aligned}
b_n &= \frac{1}{3}\int_{-3}^{0} -t\sin\frac{n\pi t}{3}\,dt + \frac{1}{3}\int_{0}^{3} t\sin\frac{n\pi t}{3}\,dt \\
&= -\frac{1}{3}\left[\frac{9}{n^2\pi^2}\sin\frac{n\pi t}{3} - \frac{3t}{n\pi}\cos\frac{n\pi t}{3}\right]_{-3}^{0} + \frac{1}{3}\left[\frac{9}{n^2\pi^2}\sin\frac{n\pi t}{3} - \frac{3t}{n\pi}\cos\frac{n\pi t}{3}\right]_{0}^{3} \\
&= \frac{3}{n\pi}\cos(-n\pi) - \frac{3}{n\pi}\cos n\pi = 0
\end{aligned}
$$

Substituting these coefficients into the series (1), we obtain

$$
f(t) = \frac{3}{2} - \frac{12}{\pi^2}\left(\frac{1}{1}\cos\frac{\pi t}{3} + \frac{1}{9}\cos\frac{3\pi t}{3} + \frac{1}{25}\cos\frac{5\pi t}{3} + \cdots\right)
$$

EXERCISES

Determine the Fourier expansions of the periodic functions whose definitions in one period are

1 $f(t) = \begin{cases} 1 & 0 < t < \pi/2 \\ 0 & \pi/2 < t < 2\pi \end{cases}$

2 $f(t) = \begin{cases} 0 & -\pi < t < 0 \\ t & 0 < t < \pi \end{cases}$

3 $f(t) = \sin t/2 \qquad -\pi < t < \pi$

4 $f(t) = \cos t \qquad -\pi/2 < t < \pi/2$

5 $f(t) = \begin{cases} 2 & 0 < t < 2\pi/3 \\ 1 & 2\pi/3 < t < 4\pi/3 \\ 0 & 4\pi/3 < t < 2\pi \end{cases}$

6 $f(t) = \begin{cases} 0 & -3 < t < -1 \\ 1 + \cos\pi t & -1 < t < 1 \\ 0 & 1 < t < 3 \end{cases}$

7 $f(t) = t \qquad -\pi < t < \pi$

8 $f(t) = e^{-t} \qquad 0 < t < 1$

9 $f(t) = \pi^2 - t^2 \qquad -\pi < t < \pi$

10 $f(t) = t - t^3 \qquad -1 < t < 1$

11 $f(t) = \begin{cases} 0 & -2 < t < -1 \\ 1 & -1 < t < 0 \\ -1 & 0 < t < 1 \\ 0 & 1 < t < 2 \end{cases}$

12 $f(t) = \begin{cases} 0 & -2 < t < -1 \\ 1 + t & -1 < t < 0 \\ 1 - t & 0 < t < 1 \\ 0 & 1 < t < 2 \end{cases}$

13 $f(t) = \begin{cases} \cos t & -\pi < t < 0 \\ \sin t & 0 < t < \pi \end{cases}$

14 $f(t) = \begin{cases} 0 & -\pi < t < 0 \\ t^2 & 0 < t < \pi \end{cases}$

15 Establish the following numerical results

$$
1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \cdots = \frac{\pi^2}{6}
$$

$$
1 - \frac{1}{2^2} + \frac{1}{3^2} - \frac{1}{4^2} + \frac{1}{5^2} - \cdots = \frac{\pi^2}{12}
$$

$$
1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \frac{1}{9^2} + \cdots = \frac{\pi^2}{8}
$$

(**Hint** Use the results of Exercise 14)

6.3

Half-range expansions

When $f(t)$ possesses certain symmetry properties, the coefficients in its Fourier expansion become especially simple This was illustrated in Example 2 of the last section, where the given function was symmetric in the y-axis and its expansion contained only cosine terms, i e , only terms which themselves were symmetric in the vertical axis In this section we shall investigate in detail just what effect the symmetry of $f(t)$ has on the coefficients in the Fourier series for $f(t)$

Suppose first that $f(t)$ is an **even** function, i e , suppose that

$$f(-t) = f(t) \qquad \text{for all } t$$

or, geometrically, that the graph of $f(t)$ is symmetric in the vertical axis Taking $d = -p$ in the formula for a_n, Eq (10), Sec 6 2, we can write

$$a_n = \frac{1}{p}\int_{-p}^{p} f(t)\cos\frac{n\pi t}{p}\,dt = \frac{1}{p}\int_{-p}^{0} f(t)\cos\frac{n\pi t}{p}\,dt + \frac{1}{p}\int_{0}^{p} f(t)\cos\frac{n\pi t}{p}\,dt$$

Now, in the integral from $-p$ to 0, let us make the substitution

$$t = - \qquad dt = -ds$$

Then, since $t = -p$ implies $s = p$ and $t = 0$ implies $s = 0$, the integral becomes

$$(1) \qquad \frac{1}{p}\int_{p}^{0} f(-s)\cos\frac{-n\pi s}{p}\,(-ds)$$

But $f(-s) = f(s)$, from the hypothesis that $f(t)$ is an even function Moreover, the cosine is also an even function, that is,

$$\cos\frac{-n\pi s}{p} = \cos\frac{n\pi s}{p}$$

Finally, the negative sign associated with ds in (1) can be eliminated by changing the limits back to the normal order, 0 to p The integral (1) then becomes

$$\frac{1}{p}\int_{0}^{p} f(s)\cos\frac{n\pi s}{p}\,ds$$

and thus a_n can be written

$$\begin{aligned} a_n &= \frac{1}{p}\int_{0}^{p} f(s)\cos\frac{n\pi s}{p}\,ds + \frac{1}{p}\int_{0}^{p} f(t)\cos\frac{n\pi t}{p}\,dt \\ &= \frac{2}{p}\int_{0}^{p} f(t)\cos\frac{n\pi t}{p}\,dt \end{aligned}$$

since the two integrals are identical, except for the dummy variable of integration, which is immaterial

Similarly, we can write

$$b_n = \frac{1}{p}\int_{-p}^{0} f(t)\sin\frac{n\pi t}{p}\,dt + \frac{1}{p}\int_{0}^{p} f(t)\sin\frac{n\pi t}{p}\,dt$$

Again, putting $t = -s$ and $dt = -ds$

in the first integral, we find

$$b_n = \frac{1}{p}\int_p^0 f(-s)\sin\frac{-n\pi s}{p}(-ds) + \frac{1}{p}\int_0^p f(t)\sin\frac{n\pi t}{p}\,dt$$

But, by hypothesis, $f(-s) = f(s)$

and $\sin\frac{-n\pi s}{p} = -\sin\frac{n\pi s}{p}$

Hence, reversing the limits on the first integral, as before, we have

$$b_n = -\frac{1}{p}\int_0^p f(s)\sin\frac{n\pi s}{p}\,ds + \frac{1}{p}\int_0^p f(t)\sin\frac{n\pi t}{p}\,dt = 0$$

since, except for the irrelevant variable of integration, the two integrals are identical in all but sign Thus we have established the following useful result

THEOREM 1

If $f(t)$ is an even periodic function, then the coefficients in the Fourier series of $f(t)$ are given by the formulas

$$a_n = \frac{2}{p}\int_0^p f(t)\cos\frac{n\pi t}{p}\,dt \qquad b_n \equiv 0$$

Now suppose that $f(t)$ is an **odd function**; i e , suppose that

$f(-t) = -f(t)$ for all t

or, geometrically, that the graph of $f(t)$ is symmetric in the origin Then proceeding just as before, we can write

$$\begin{aligned} a_n &= \frac{1}{p}\int_{-p}^0 f(t)\cos\frac{n\pi t}{p}\,dt + \frac{1}{p}\int_0^p f(t)\cos\frac{n\pi t}{p}\,dt \\ &= \frac{1}{p}\int_p^0 f(-s)\cos\frac{-n\pi s}{p}(-ds) + \frac{1}{p}\int_0^p f(t)\cos\frac{n\pi t}{p}\,dt \\ &= -\frac{1}{p}\int_0^p f(s)\cos\frac{n\pi s}{p}\,ds + \frac{1}{p}\int_0^p f(t)\cos\frac{n\pi t}{p}\,dt \\ &= 0 \end{aligned}$$

$$\begin{aligned} b_n &= \frac{1}{p}\int_{-p}^0 f(t)\sin\frac{n\pi t}{p}\,dt + \frac{1}{p}\int_0^p f(t)\sin\frac{n\pi t}{p}\,dt \\ &= \frac{1}{p}\int_p^0 f(-s)\sin\frac{-n\pi s}{p}(-ds) + \frac{1}{p}\int_0^p f(t)\sin\frac{n\pi t}{p}\,dt \\ &= \frac{1}{p}\int_0^p f(s)\sin\frac{n\pi s}{p}\,ds + \frac{1}{p}\int_0^p f(t)\sin\frac{n\pi t}{p}\,dt \\ &= \frac{2}{p}\int_0^p f(t)\sin\frac{n\pi t}{p}\,dt \end{aligned}$$

Thus we have established the following result

THEOREM 2

If $f(t)$ is an odd periodic function, then the coefficients in the Fourier series of $f(t)$ are given by the formulas

$$a_n \equiv 0 \qquad b_n = \frac{2}{p}\int_0^p f(t)\sin\frac{n\pi t}{p}\,dt$$

It should be emphasized here that oddness and evenness are not intrinsic properties of a graph, but depend upon its relation to the axes of the coordinate system. For instance, in Fig. 6.4, if the line AA' is chosen as the vertical axis, the graph represents an even function whose Fourier expansion, in accordance with Theorem 1, will contain only cosine terms. On the other hand, if BB' is chosen as the vertical axis, the graph represents an odd function, and, by Theorem 2, only sine terms will appear in its expansion. Finally, if a general line, such as CC', is chosen as the vertical axis, the graph represents a function which is neither odd nor even, and both sines and cosines will appear in its Fourier series.

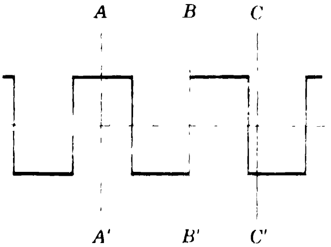


FIGURE 6.4 Plot showing how oddness and evenness depend on the choice of axes

The observations we have just made about the Fourier coefficients of odd and even functions serve to reduce by half the labor of expanding such functions. However, their chief value is that they allow us to meet the requirements of certain problems* in which expansions containing *only* cosine terms or expansions containing *only* sine terms must be constructed.

Let us suppose that the conditions of a problem require us to consider the values of a function *only* in the interval from 0 to p. In other words, conditions of periodicity are irrelevant to the problem, and what the function may be outside the range $(0,p)$ is completely immaterial. This being the case, we can define the function in any way we please over the interval $(-p,0)$ and *then* use the Euler formulas to determine the coefficients in the Fourier series of its periodic extension. Between $-p$ and 0 this series will, of course, converge to whatever extension we created over this interval, but *irrespective of this extension* the series will represent the given function between 0 and p, as required.

In particular, if we extend the function from 0 to $-p$

---

* Examples of such problems will be found in Sec. 8.4.

by reflecting it in the vertical axis, so that $f(-t) = f(t)$, the original function together with its extension is even; hence, the Fourier expansion of its periodic continuation will contain only cosine terms [including, of course, the constant term $a_0/2 \equiv (a_0/2)\cos(0\pi t/p)$] whose coefficients, as we showed above, will be given by

$$a_n = \frac{2}{p}\int_0^p f(t)\cos\frac{n\pi t}{p}\,dt$$

On the other hand, if we extend the function from 0 to $-p$ by reflecting it in the origin, so that $f(-t) = -f(t)$, the extended function is odd, and hence the Fourier series of its periodic continuation will contain only sine terms, whose coefficients will be given by

$$b_n = \frac{2}{p}\int_0^p f(t)\sin\frac{n\pi t}{p}\,dt$$

Thus, simply by imagining the appropriate extension of a function originally defined only for $0 < t < p$, we can obtain expansions representing the function on this interval and containing only cosine terms or only sine terms, as we please Such series are known as **half-range expansions.**

EXAMPLE 1

Find the half-range expansions of the function

$$f(t) = t - t^2 \qquad 0 < t < 1$$

The half-range cosine expansion is obtained by first extending $t - t^2$ from the given interval $(0,1)$ to the interval $(-1,0)$ by reflection in the y-axis and then taking the function thus defined from -1 to 1 as one period of a periodic function of period $2p = 2$ However, once we understand the reasoning underlying the procedure we need give no thought to the extension but can write immediately, on the basis of Theorem 1,

$$b_n = 0$$

and

$$\begin{aligned} a_n &= \frac{2}{1}\int_0^1 (t - t^2)\cos\frac{n\pi t}{1}\,dt \\ &= 2\left[\left(\frac{\cos n\pi t}{n^2\pi^2} + \frac{t}{n\pi}\sin n\pi t\right) - \left(\frac{2t}{n^2\pi^2}\cos n\pi t - \frac{2}{n^3\pi^3}\sin n\pi t + \frac{t^2}{n\pi}\sin n\pi t\right)\right]_0^1 \\ &= 2\left(\frac{\cos n\pi - 1}{n^2\pi^2} - \frac{2\cos n\pi}{n^2\pi^2}\right) \\ &= -\frac{2(1 + \cos n\pi)}{n^2\pi^2} \qquad n \neq 0 \end{aligned}$$

$$a_0 = \frac{2}{1}\int_0^1 (t - t^2)\,dt = 2\left[\frac{t^2}{2} - \frac{t^3}{3}\right]_0^1 = \frac{1}{3}$$

Hence it is possible to represent $f(t) = t - t^2$ for $0 < t < 1$ by the series

$$f(t) = \frac{1}{6} - \frac{4}{\pi^2}\left(\frac{\cos 2\pi t}{4} + \frac{\cos 4\pi t}{16} + \frac{\cos 6\pi t}{36} + \frac{\cos 8\pi t}{64} + \cdots\right) \tag{2}$$

Similarly, the half-range sine expansion is obtained by first extending the given function $t - t^2$ to the interval $(-1,0)$ by reflection in the origin and then extending periodically the function thus defined over $(-1,1)$. However, all we need to obtain the expansion is to note that, according to Theorem 2,

$$a_n = 0$$

and
$$
\begin{aligned}
b_n &= \frac{2}{1}\int_0^1 (t - t^2)\sin\frac{n\pi t}{1}\,dt \\
&= 2\left[\left(\frac{1}{n^2\pi^2}\sin n\pi t - \frac{t}{n\pi}\cos n\pi t\right) - \left(\frac{2t}{n^2\pi^2}\sin n\pi t + \frac{2}{n^3\pi^3}\cos n\pi t - \frac{t^2}{n\pi}\cos n\pi t\right)\right]_0^1 \\
&= 2\left[\left(-\frac{\cos n\pi}{n\pi}\right) - \left(\frac{2(\cos n\pi - 1)}{n^3\pi^3} - \frac{\cos n\pi}{n\pi}\right)\right] \\
&= \frac{4(1 - \cos n\pi)}{n^3\pi^3}
\end{aligned}
$$

Hence it is also possible to represent $f(t)$ for $0 < t < 1$ by the series

$$f(t) = \frac{8}{\pi^3}\left(\frac{\sin \pi t}{1} + \frac{\sin 3\pi t}{27} + \frac{\sin 5\pi t}{125} + \frac{\sin 7\pi t}{343} + \cdots\right) \tag{3}$$

Series (2) and (3) are by no means the only Fourier series that will represent $t - t^2$ on the interval $(0,1)$. They are merely the most convenient or most useful ones. In fact, with every possible extension of $t - t^2$ from 0 to -1 there is associated a series yielding $t - t^2$ for $0 < t < 1$. For instance, a third such series might be obtained by letting the extension be simply the function defined by $t - t^2$ itself for $-1 < t < 0$. In this case

$$
\begin{aligned}
a_n &= \frac{1}{1}\int_{-1}^1 (t - t^2)\cos\frac{n\pi t}{1}\,dt \\
&= \left[\left(\frac{\cos n\pi t}{n^2\pi^2} + \frac{t}{n\pi}\sin n\pi t\right) - \left(\frac{2t}{n^2\pi^2}\cos n\pi t - \frac{2}{n^3\pi^3}\sin n\pi t + \frac{t^2}{n\pi}\sin n\pi t\right)\right]_{-1}^1 \\
&= -\frac{4\cos n\pi}{n^2\pi^2} \qquad n \neq 0
\end{aligned}
$$

$$a_0 = \frac{1}{1}\int_{-1}^1 (t - t^2)\,dt = \left[\frac{t^2}{2} - \frac{t^3}{3}\right]_{-1}^1 = -\frac{2}{3}$$

$$
\begin{aligned}
b_n &= \frac{1}{1}\int_{-1}^1 (t - t^2)\sin\frac{n\pi t}{1}\,dt \\
&= \left[\left(\frac{1}{n^2\pi^2}\sin n\pi t - \frac{t}{n\pi}\cos n\pi t\right) - \left(\frac{2t}{n^2\pi^2}\sin n\pi t + \frac{2}{n^3\pi^3}\cos n\pi t - \frac{t^2}{n\pi}\cos n\pi t\right)\right]_{-1}^1 \\
&= -\frac{2\cos n\pi}{n\pi}
\end{aligned}
$$

Hence, for $0 < t < 1$ it is also possible to write

$$
\begin{aligned}
f(t) = -\frac{1}{3} &+ \frac{4}{\pi^2}\left(\frac{\cos \pi t}{1} - \frac{\cos 2\pi t}{4} + \frac{\cos 3\pi t}{9} - \frac{\cos 4\pi t}{16} + \cdots\right) \\
&+ \frac{2}{\pi}\left(\frac{\sin \pi t}{1} - \frac{\sin 2\pi t}{2} + \frac{\sin 3\pi t}{3} - \frac{\sin 4\pi t}{4} + \cdots\right)
\end{aligned} \tag{4}
$$

FIGURE 6.5
Plot showing different periodic functions coinciding over the interval (0,1)

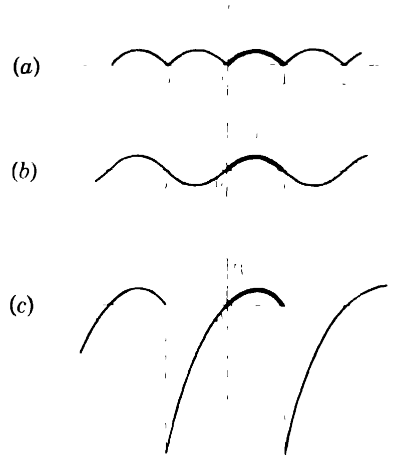


Figure 6.5*a*, *b*, and *c* shows the extended periodic functions represented, respectively, by the series (2), (3), and (4)

Figure 6.5 and the associated expansions illustrate another interesting and important fact In Fig. 6.5*c* the graph as a whole is not continuous but has jumps at $t = \pm 1, \pm 3, \pm 5, \ldots$ In the corresponding series (4), the coefficients (of the sine terms) decrease only at a rate proportional to $1/n$ On the other hand, the graph in Fig 6.5*a* is everywhere continuous but has corners, or points where the tangent changes direction discontinuously In the corresponding series (2), the coefficients all become small much more rapidly than in (4), in fact, they decrease at a rate proportional to $1/n^2$ Finally, the graph in Fig 6.5*b* not only is continuous but has a continuous tangent, i e , there are no points where the tangent changes direction abruptly This smoother behavior of the function is reflected in the coefficients in the corresponding series (3), which in this case approach zero at a rate proportional to $1/n^3$ These observations are summed up and generalized in the following theorem, which we cite without proof *

THEOREM 3

As n becomes infinite, the coefficients a_n and b_n in the Fourier expansion of a periodic function satisfying the Dirichlet conditions always approach zero at least as rapidly as c/n, where c is a constant independent of n If the function has one or more points of discontinuity, then either a_n or b_n, and in general both, can decrease no faster than this In general, if a function $f(t)$ and its first $k - 1$ derivatives satisfy the Dirichlet conditions and are everywhere continuous, then as n

---

* See, for instance, H. S Carslaw, "Fourier Series," pp 269-271, Dover Publications, Inc., New York, 1930.

becomes infinite the coefficients a_n and b_n in the Fourier series of $f(t)$ tend to zero at least as rapidly as c/n^{k+1} If, in addition, the kth derivative of $f(t)$ is not everywhere continuous, then either a_n or b_n, and in general both, can tend to zero no faster than c/n^{k+1}

More concisely, though less accurately, this theorem asserts that the smoother the function, the faster its Fourier expansion converges

Closely associated with the last result are the following observations, which we also state without proof

THEOREM 4*

The integral of any periodic function which satisfies the Dirichlet conditions can be found by term-by-term integration of the Fourier series of the function

THEOREM 5†

If $f(t)$ is a periodic function which satisfies the Dirichlet conditions and is everywhere continuous and if $f'(t)$ also satisfies the Dirichlet conditions, then wherever it exists, $f'(t)$ can be found by term-by-term differentiation of the Fourier series of $f(t)$

EXERCISES

1 By considering the identity $f(t) = \frac{1}{2}[f(t) + f(-t)] + \frac{1}{2}[f(t) - f(-t)]$, show that any function, defined for both positive and negative values of t, can be written as the sum of an even function and an odd function

Obtain the half-range sine and cosine expansions of each of the following functions

2 $f(t) = 1 \qquad 0 < t < 1$

3 $f(t) = e^t \qquad 0 < t < 1$

4 $f(t) = \cos t \qquad 0 < t < 2\pi$

5 $f(t) = \sin t \qquad 0 < t < 2\pi$

6 $f(t) = \begin{cases} t & 0 < t < 2 \\ 6 - 2t & 2 < t < 3 \end{cases}$

7 $f(t) = \begin{cases} t^2 & 0 < t < 1 \\ 2 - t & 1 < t < 2 \end{cases}$

8 Obtain a series, different from the half-range sine expansion, which will represent $t - t^2$ for $0 < t < 1$ and whose coefficients will decrease as $1/n^3$

9 Is it possible to obtain a series representing $t - t^2$ for $0 < t < 1$ whose coefficients will decrease as $1/n^4$?

10 Find a function whose half-range cosine series will have coefficients decreasing as $1/n^4$ Determine the expansion

11 How rapidly will the coefficients in the Fourier series of the periodic function $1/(2 + \cos t)$ decrease?

12 If $f(t) = \begin{cases} 1 & 0 < t < a \\ \dfrac{t-1}{a-1} & a < t < 1 \\ 0 & 1 < t < 2 \end{cases}$ and if a is only slightly less than 1, discuss the behavior of the coefficients in the half-range cosine expansion of $f(t)$ for small and medium values of n as well as for $n \to \infty$

---

* See, for instance, E C Titchmarsh, "Theory of Functions," pp. 419–421, Oxford Book Company, Inc , New York, 1939

† See, for instance, E T Whittaker and G N Watson, "Modern Analysis," pp. 168–169, The Macmillan Company, New York, 1943

13 If $f(t)$, originally defined only for $0 < t < p$, is extended from p to $2p$ by reflection in the line $t = p$, show that the half-range sine expansion of the extended function contains no terms of the form

$$\sin \frac{n\pi t}{2p} \qquad n \text{ even}$$

and show that the coefficients of the terms of the form

$$\sin \frac{n\pi t}{2p} \qquad n \text{ odd}$$

are given by the formula

$$b_n = \frac{2}{p} \int_0^p f(t) \sin \frac{n\pi t}{2p}\, dt$$

14 Determine how $f(t)$, originally defined only for $0 < t < p$, must be extended from p to $2p$ in order that the half-range cosine expansion of the extended function will contain no terms of the form

$$\cos \frac{n\pi t}{2p} \qquad n \text{ even}$$

Derive a formula for the nonzero coefficients

15 Prove Theorem 3 for the special case $k = 1$ Under what conditions can either a_n or b_n decrease faster than c/n^2? [Hint Assuming that $f'(t)$ has a single point of discontinuity in each period, apply integration by parts, with $f(t) = u$, to the integrals defining a_n and b_n]

6.4

Alternative forms of Fourier series

The original form of the Fourier series of a function, as derived in Sec 6 2, can be converted into several other trigonometric forms and into one in which imaginary exponentials appear instead of real trigonometric functions For instance, in the series

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi t}{p} + b_n \sin \frac{n\pi t}{p} \right)$$

we can apply to each pair of terms of the same frequency the usual procedure for reducing the sum of a sine and a cosine of the same angle to a single term

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \sqrt{a_n^2 + b_n^2} \left(\frac{a_n}{\sqrt{a_n^2 + b_n^2}} \cos \frac{n\pi t}{p} + \frac{b_n}{\sqrt{a_n^2 + b_n^2}} \sin \frac{n\pi t}{p} \right)$$

If we now define the angles γ_n and δ_n from the triangle shown in Fig 6 6 and set

$$A_0 = \frac{a_0}{2} \qquad \text{and} \qquad A_n = \sqrt{a_n^2 + b_n^2}$$

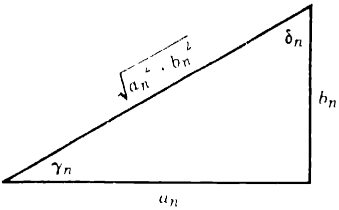


FIGURE 6.6 The triangle defining the phase angles γ_n and δ_n for the resultant of the terms of frequency $n\pi/p$ in a Fourier series

the last series can be written

$$f(t) = A_0 + \sum_{n=1}^{\infty} A_n \left(\cos \frac{n\pi t}{p} \cos \gamma_n + \sin \frac{n\pi t}{p} \sin \gamma_n \right)$$

$$= A_0 + \sum_{n=1}^{\infty} A_n \cos \left(\frac{n\pi t}{p} - \gamma_n \right)$$

or, equally well,

$$f(t) = A_0 + \sum_{n=1}^{\infty} A_n \left(\cos \frac{n\pi t}{p} \sin \delta_n + \sin \frac{n\pi t}{p} \cos \delta_n \right)$$

$$= A_0 + \sum_{n=1}^{\infty} A_n \sin \left(\frac{n\pi t}{p} + \delta_n \right)$$

In either of these forms, the quantity $A_n = \sqrt{a_n^2 + b_n^2}$ is the resultant amplitude of the components of frequency $n\pi/p$, that is, the amplitude of the nth harmonic in the expansion. The phase angles

$$\gamma_n = \tan^{-1} \frac{b_n}{a_n} \qquad \text{and} \qquad \delta_n = \tan^{-1} \frac{a_n}{b_n} = \frac{\pi}{2} - \gamma_n$$

measure the lag or lead of the nth harmonic with reference to a pure cosine or pure sine wave of the same frequency.

The complex exponential form of a Fourier series is obtained by substituting the exponential equivalents of the cosine and sine terms into the original form of the series

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \frac{e^{in\pi t/p} + e^{-in\pi t/p}}{2} + b_n \frac{e^{in\pi t/p} - e^{-in\pi t/p}}{2i} \right)$$

Collecting terms on the various exponentials and noting that $1/i = -i$, we obtain

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(\frac{a_n - ib_n}{2} e^{in\pi t/p} + \frac{a_n + ib_n}{2} e^{-in\pi t/p} \right)$$

If we now define

$$c_0 = \frac{a_0}{2} \qquad c_n = \frac{a_n - ib_n}{2} \qquad c_{-n} = \frac{a_n + ib_n}{2}$$

the last series can be written in the more symmetric form

(1) $$f(t) = \sum_{n=-\infty}^{n=\infty} c_n e^{ni\pi t/p}$$

Now, when it is used at all, this exponential form is used as a basic form in its own right, i.e., it is not obtained by transformation from the trigonometric form, but is constructed directly from the given function. To do this requires that expressions be available for the direct evaluation of the coefficients c_n. These may easily be found from the definitions of c_0, c_n, and c_{-n}. For

$$c_0 = \frac{1}{2} a_0 = \frac{1}{2p} \int_d^{d+2p} f(t)\, dt$$

$$\begin{aligned} c_n &= \frac{a_n - ib_n}{2} = \frac{1}{2}\left[\frac{1}{p}\int_d^{d+2p} f(t) \cos\frac{n\pi t}{p}\, dt - i\,\frac{1}{p}\int_d^{d+2p} f(t) \sin\frac{n\pi t}{p}\, dt\right] \\ &= \frac{1}{2p}\int_d^{d+2p} f(t)\left(\cos\frac{n\pi t}{p} - i\sin\frac{n\pi t}{p}\right) dt \\ &= \frac{1}{2p}\int_d^{d+2p} f(t) e^{-ni\pi t/p}\, dt \end{aligned}$$

$$\begin{aligned} c_{-n} &= \frac{a_n + ib_n}{2} = \frac{1}{2}\left[\frac{1}{p}\int_d^{d+2p} f(t) \cos\frac{n\pi t}{p}\, dt + i\,\frac{1}{p}\int_d^{d+2p} f(t) \sin\frac{n\pi t}{p}\, dt\right] \\ &= \frac{1}{2p}\int_d^{d+2p} f(t)\left(\cos\frac{n\pi t}{p} + i\sin\frac{n\pi t}{p}\right) dt \\ & \frac{1}{2p}\int_d^{d+2p} f(t) e^{ni\pi t/p}\, dt \end{aligned}$$

Clearly, whether the index n is positive, negative, or zero, c_n is given by the one formula

(2) $$c_n = \frac{1}{2p}\int_d^{d+2p} f(t) e^{-ni\pi t/p}\, dt$$

As usual, d will almost always be either $-p$ or 0.

In the complex representation defined by (1) and (2), a certain symmetry between the expressions for a function and for its Fourier coefficients is evident. In fact the expressions

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{ni\pi t/p}$$

$$c_n = \frac{1}{2p}\int_{-p}^{p} f(t) e^{-ni\pi t/p}\, dt$$

are of essentially the same structure, as the following correlation reveals

$$t \sim n$$

$$f(t) \sim c_n \equiv c(n)$$

$$e^{ni\pi t/p} \sim e^{-ni\pi t/p}$$

$$\sum_{n=-\infty}^{\infty} (\quad) \sim \frac{1}{2p}\int_{-p}^{p} (\quad)\, dt$$

This duality is worthy of note, and, as our development proceeds to the Fourier integral and the Laplace transform, it will become still more striking and fundamental.

EXAMPLE 1

Find the complex form of the Fourier series of the function whose definition in one period is $f(t) = e^{-t}$, $-1 < t < 1$

Since $p = 1$, we have from (2), taking $d = -1$,

$$c_n = \frac{1}{2}\int_{-1}^{1} e^{-t}e^{-ni\pi t}\,dt = \frac{1}{2}\left[\frac{e^{-(1+ni\pi)t}}{-(1+ni\pi)}\right]_{-1}^{1}$$

$$= \frac{e^{-(1+ni\pi)} - e^{(1+ni\pi)}}{-2(1+ni\pi)}$$

$$= \frac{e\,e^{ni\pi} - e^{-1}\,e^{-ni\pi}}{2(1+ni\pi)}$$

Now $e^{i\pi} = \cos\pi + i\sin\pi = -1$, and thus $e^{ni\pi} = e^{-ni\pi} = (-1)^n$. Hence

$$c_n = \frac{(-1)^n}{(1+ni\pi)}\,\frac{e - e^{-1}}{2} = \frac{(-1)^n(1 - ni\pi)\sinh 1}{1+n^2\pi^2}$$

The expansion of $f(t)$ is therefore

$$f(t) = \sum_{n=-\infty}^{\infty} (-1)^n \frac{(1 - ni\pi)\sinh 1}{1+n^2\pi^2}\, e^{ni\pi t}$$

This, of course, can be converted into the real trigonometric form without difficulty, for we have, by definition,

$$c_n = \frac{a_n - ib_n}{2}$$

$$c_{-n} = \frac{a_n + ib_n}{2}$$

and thus, by adding and subtracting,

$$a_n = c_n + c_{-n} \qquad b_n = i(c_n - c_{-n})$$

Therefore in this problem

$$a_n = \frac{(-1)^n(1 - ni\pi)\sinh 1}{1+n^2\pi^2} + \frac{(-1)^n(1 + ni\pi)\sinh 1}{1+n^2\pi^2} = \frac{(-1)^n\,2\sinh 1}{1+n^2\pi^2}$$

$$b_n = i\left[\frac{(-1)^n(1 - ni\pi)\sinh 1}{1+n^2\pi^2} - \frac{(-1)^n(1 + ni\pi)\sinh 1}{1+n^2\pi^2}\right] = \frac{(-1)^n\,2n\pi\sinh 1}{1+n^2\pi^2}$$

$$\tfrac{1}{2}a_0 = c_0 = \sinh 1$$

Hence, we can also write

$$f(t) = \sinh 1 - 2\sinh 1\left(\frac{\cos\pi t}{1+\pi^2} - \frac{\cos 2\pi t}{1+4\pi^2} + \frac{\cos 3\pi t}{1+9\pi^2} - \cdots\right)$$

$$- 2\pi\sinh 1\left(\frac{\sin\pi t}{1+\pi^2} - \frac{2\sin 2\pi t}{1+4\pi^2} + \frac{3\sin 3\pi t}{1+9\pi^2} - \cdots\right)$$

EXERCISES

What is the amplitude of the resultant term of frequency $n\pi/p$ in the Fourier series of the functions whose definitions in one period are the following? What is the phase of each of these terms relative to $\cos n\pi t/p$? relative to $\sin n\pi t/p$?

1 $f(t) = \begin{cases} 1 & 0 < t < 1 \\ 0 & 1 < t < 4 \end{cases}$

2 $f(t) = \begin{cases} t & 0 < t < 1 \\ 0 & 1 < t < 2 \end{cases}$

3 $f(t) = \begin{cases} 1 & 0 < t < 1 \\ -1 & 1 < t < 2 \\ 0 & 2 < t < 4 \end{cases}$

4 $f(t) = \begin{cases} 1 & 0 < t < 1 \\ 0 & 1 < t < 2 \\ -1 & 2 < t < 3 \\ 0 & 3 < t < 4 \end{cases}$

Find the complex form of the Fourier series of the periodic functions whose definitions in one period are

5 $f(t) = \begin{cases} 1 & 0 < t < 1 \\ 0 & 1 < t < 2 \end{cases}$

6 $f(t) = \begin{cases} 1 & 0 < t < 1 \\ -1 & 1 < t < 2 \end{cases}$

7 $f(t) = t \quad 0 < t < 1$

8 $f(t) = t \quad -1 < t < 1$

9 $f(t) = \cos t \quad -\pi/2 < t < \pi/2$ [Hint: Use the fact that $\cos\theta = \frac{1}{2}(e^{i\theta} + e^{-i\theta})$.]

10 $f(t) = \sin t \quad 0 < t < \pi$

6.5

Applications

Although we shall see other uses of Fourier series in later chapters, their most important application at the present stage of our work is in the analysis of the behavior of physical systems subjected to periodic disturbances.

EXAMPLE 1

If the **root-mean-square, or rms, value** of a function $f(t)$ over an interval (a,b) is defined as

$$\sqrt{\frac{\int_a^b f^2(t)\,dt}{b-a}} \tag{1}$$

express the rms value of a periodic function over one period in terms of the coefficients in its Fourier expansion.

If $f(t)$ is of period $2p$, we can write

$$f(t) = \frac{a_0}{2} + a_1\cos\frac{\pi t}{p} + \cdots + a_n\cos\frac{n\pi t}{p} + \cdots + b_1\sin\frac{\pi t}{p} + \cdots + b_n\sin\frac{n\pi t}{p} + \cdots$$

Hence, $f^2(t)$ will consist exclusively of squared terms of the form

$$\frac{a_0^2}{4} \qquad a_n^2\cos^2\frac{n\pi t}{p} \qquad b_n^2\sin^2\frac{n\pi t}{p}$$

and cross-product terms of the form

$$a_0a_n\cos\frac{n\pi t}{p} \qquad a_0b_n\sin\frac{n\pi t}{p}$$

$$2a_ma_n\cos\frac{m\pi t}{p}\cos\frac{n\pi t}{p} \qquad 2a_mb_n\cos\frac{m\pi t}{p}\sin\frac{n\pi t}{p} \qquad 2b_mb_n\sin\frac{m\pi t}{p}\sin\frac{n\pi t}{p}$$

As in the original derivation of the Euler formulas in Sec. 6.2, the integral of every cross-product term, taken over one period of the function, is zero. Moreover, for the squared terms we have

$$\frac{a_0^2}{4}\int_{-p}^{p} dt = \frac{a_0^2 p}{2} \qquad a_n^2\int_{-p}^{p}\cos^2\frac{n\pi t}{p}\,dt = a_n^2 p \qquad b_n^2\int_{-p}^{p}\sin^2\frac{n\pi t}{p}\,dt = b_n^2 p$$

Hence, dividing each of the nonzero terms by the length of the period, $2p$, we obtain for the required rms value

$$f(t)\Big|_{\text{rms}} = \sqrt{\frac{a_0^2}{4} + \frac{1}{2}\sum_{n=1}^{\infty}(a_n^2 + b_n^2)} \tag{2}$$

Since the coefficients in the complex exponential form of the Fourier series of $f(t)$ are related to the coefficients in the real trigonometric form by the equations

$$c_0 = \frac{a_0}{2} \qquad c_n = \frac{a_n - ib_n}{2} \qquad c_{-n} = \frac{a_n + ib_n}{2} = \bar{c}_n$$

Eq. (2) can also be written

$$f(t)\Big|_{\text{rms}} = \sqrt{c_0^2 + 2\sum_{n=1}^{\infty} c_n\bar{c}_n} \tag{3}$$

In particular, if $i = f(t)$ is an electric current flowing through a resistance R, the average power dissipated is

$$i_{\text{rms}}^2 R = \left[\frac{a_0^2}{4} + \frac{1}{2}\sum_{n=1}^{\infty}(a_n^2 + b_n^2)\right]R = \left(c_0^2 + 2\sum_{n=1}^{\infty} c_n\bar{c}_n\right)R$$

EXAMPLE 2

Determine the steady-state forced vibrations of the system shown in Fig. 6.7a if the applied force is as shown in Fig. 6.7b.

Our first step must be to obtain the Fourier expansion of the driving force. Since $F(t)$ is clearly an odd function of t, no cosine terms can be present, and thus we need only compute b_n

$$\begin{aligned} b_n &= \frac{2}{\frac{1}{2}}\int_0^{1/2} 20\sin\frac{n\pi t}{\frac{1}{2}}\,dt = 80\left[-\frac{\cos 2n\pi t}{2n\pi}\right]_0^{1/2} \\ &= 40\,\frac{1-\cos n\pi}{n\pi} \\ &= \begin{cases} 0 & n \text{ even} \\ \dfrac{80}{n\pi} & n \text{ odd} \end{cases} \end{aligned}$$

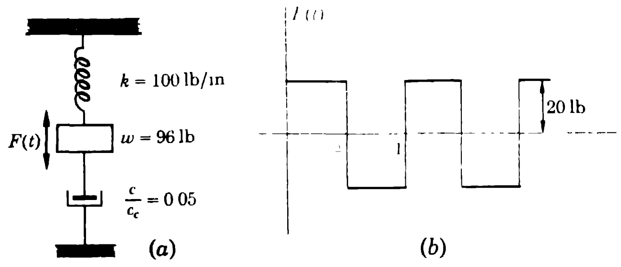


FIGURE 6.7 A spring-mass system acted upon by an alternating square-wave force.

Hence

$$F(t) = \frac{80}{\pi}\left(\sin 2\pi t + \frac{\sin 6\pi t}{3} + \frac{\sin 10\pi t}{5} + \frac{\sin 14\pi t}{7} + \cdots\right)$$

Since we are concerned only with the steady-state forced motion of the system, we need determine only the particular integral corresponding to $F(t)$. Since the equation is linear, this can be done very simply using the ideas of Sec. 5.3, for it is necessary only to apply the proper magnification ratio and phase shift to each component of the driving force and add the results. Preparatory to this, we must determine the static deflections that would be produced in the system by steady forces having the magnitudes of the various terms of $F(t)$. These are given by

$$(\delta_{st})_n = \frac{(80/n\pi)\ \text{lb}}{100\ \text{lb/in}} = \frac{4}{5n\pi}\ \text{in} \qquad n\ \text{odd}$$

Then we must calculate the undamped natural frequency of the system

$$\omega_n\dagger = \sqrt{\frac{kg}{w}} = \sqrt{\frac{100 \times 384}{96}} = 20\ \text{rad/sec}$$

The rest of the work can best be presented in tabular form

Term	δ_{st}	$\dfrac{\omega}{\omega_n}$	$M = \dfrac{1}{\sqrt{\left(1 - \dfrac{\omega^2}{\omega_n^2}\right)^2 + \left(2\dfrac{c}{c_c}\dfrac{\omega}{\omega_n}\right)^2}}$	$\alpha = \tan^{-1}\dfrac{2\dfrac{c}{c_c}\dfrac{\omega}{\omega_n}}{1 - \dfrac{\omega^2}{\omega_n^2}}$	Steady-state term = $\delta_{st}M \sin(\omega t - \alpha)$
1	$\dfrac{1}{5\pi}$	$\dfrac{2\pi}{20}$	1.11	2°	$0.28 \sin(2\pi t - 2°)$
2	$\dfrac{4}{15\pi}$	$\dfrac{6\pi}{20}$	6.83	40°	$0.58 \sin(6\pi t - 40°)$
3	$\dfrac{4}{25\pi}$	$\dfrac{10\pi}{20}$	0.68	174°	$0.03 \sin(10\pi t - 174°)$
4	$\dfrac{4}{35\pi}$	$\dfrac{14\pi}{20}$	0.26	177°	$0.01 \sin(14\pi t - 177°)$

Figure 6.8 shows the steady-state displacement plotted as a function of time.

This example illustrates an exceedingly important but sometimes misunderstood characteristic of forced vibrations. If the driving force is not a pure sine or cosine function, its Fourier expansion will contain terms of frequencies above the fundamental, or apparent, frequency of the excitation. If the frequency of one of these terms happens to be close to the natural, or resonant, frequency of the system and if the amount of friction in the system is small, the corresponding magnification ratio will be large, and its value may offset many times the smaller amplitude of that harmonic and make the resultant term the dominant part of the entire response. If and when this happens, the response will appear to be of a higher frequency than the force which produces it. Figure 6.8 shows this clearly, for, although the force alternates only once per second, the weight is seen to move up and down three times per second.

It is interesting to note that, although the driving force in this example is discontinuous, both the displacement and the velocity it produces are continuous. This is suggested by the

---

† We must remember that here the subscript n in ω_n stands for *natural* and is in no way connected with the parameter n which identifies the general term in the Fourier expansion of $F(t)$. In the next section this will not be the case, but, taken in context, this dual use of the symbol ω_n should cause no confusion.

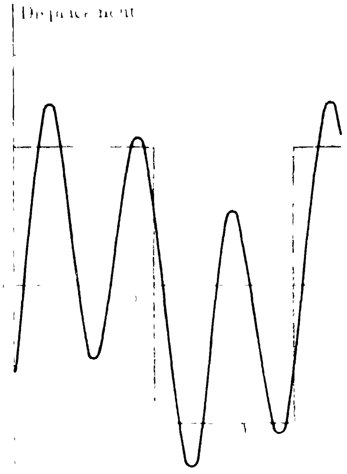


FIGURE 6.8 Plot showing a response of apparent frequency greater than that of the excitation producing it.

plot of the displacement shown in Fig 6 8 and confirmed by an application of Theorem 3, Sec 6 3 In fact, since the frequency of the nth term in the Fourier expansion of the driving force $F(t)$ is $(2n - 1)2\pi \approx \quad \pi$, it follows, neglecting all but the highest power of n, that, for n sufficiently large, the magnification ratio M is arbitrarily close to

$$\frac{1}{\omega^2/\omega_n{}^2} \approx \frac{1}{(4n\pi)^2/(20)^2} = \frac{25}{n^2\pi^2}$$

Therefore, since the static deflection corresponding to the nth term in the expansion of $F(t)$ is

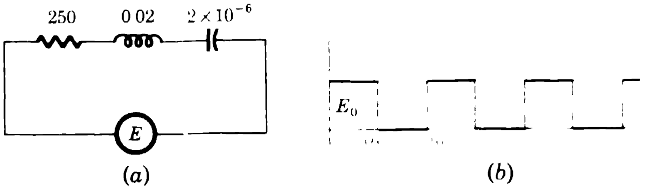
$$(\delta_{st})_n = \frac{4}{5(2n - 1)\pi} \approx \frac{2}{5n\pi}$$

it follows that as n becomes infinite the coefficient of the nth term in the expansion of the steady-state displacement, namely, $(\delta_{st})_n M$, tends to zero as $10/\pi^3 n^3$ Thus, according to Theorem 3, Sec 6 3, the displacement $y(t)$ and the velocity $y(t)$ are continuous, but the acceleration $y(t)$ is discontinuous

EXAMPLE 3

Find the steady-state current produced in the circuit shown in Fig 6 9a by the voltage shown in Fig 6 9b

Our first step must be to find the Fourier expansion of the voltage Since we plan to use the complex impedance, it will be convenient to use the complex exponential form of the Fourier



(a) (b)

FIGURE 6.9 A series circuit driven by a square-wave voltage

series. Hence we compute

$$c_n = \frac{1}{0.01}\int_0^{0.005} E_0 e^{-ni\pi t/0.005}\,dt = 100E_0 \frac{e^{-ni\pi t/0.005}}{-ni\pi/0.005}\Bigg|_0^{0.005}$$

$$= E_0 \frac{1 - e^{-ni\pi}}{2ni\pi}$$

$$= \begin{cases} 0 & n \text{ even, } n \neq 0 \\ \dfrac{E_0}{ni\pi} = -\dfrac{iE_0}{n\pi} & n \text{ odd} \end{cases}$$

$$c_0 = \frac{1}{0.01}\int_0^{0.005} E_0\,dt = \frac{E_0}{2}$$

Therefore,

$$E(t) = E_0\left(\cdots + \frac{ie^{-600i\pi t}}{3\pi} + \frac{ie^{-200i\pi t}}{\pi} + \frac{1}{2} - \frac{ie^{200i\pi t}}{\pi} - \frac{ie^{600i\pi t}}{3\pi} - \cdots\right)$$

Now, in Sec. 5.4 we showed that the steady-state current produced by a voltage of the form $Ae^{i\omega t}$ could be found simply by dividing the voltage by the complex impedance

$$Z(\omega) = R + i\left(\omega L - \frac{1}{\omega C}\right)$$

Using the data of the present problem, we have

$$Z(\omega) = 250 + i\left(0.02\omega - \frac{10^6}{2\omega}\right)$$

or, since $\omega = 200n\pi \qquad n$ odd

we have $Z(\omega) \equiv Z_n = 250 + i\left(4n\pi - \dfrac{2{,}500}{n\pi}\right) \qquad n$ odd

Hence, dividing each term in the expansion of the voltage $E(t)$ by the value of Z for the corresponding frequency, we find

$$I(t) = \sum_{n=-\infty}^{\infty} D_n e^{200ni\pi t} \qquad n \text{ odd}\dagger$$

where

$$D_n = \frac{c_n}{Z_n} = -\frac{iE_0}{n\pi}\,\frac{1}{250 + i(4n\pi - 2{,}500/n\pi)} = \frac{-iE_0}{250n\pi + i(4\pi^2n^2 - 2{,}500)} \qquad n \text{ odd}$$

If we desire the real trigonometric form of this expansion, namely,

$$I(t) = \tfrac{1}{2}a_0 + a_1\cos 200\pi t + a_3\cos 600\pi t + \cdots + b_1\sin 200\pi t + b_3\sin 600\pi t + \cdots$$

we have at once

$$a_n = D_n + D_{-n} = -iE_0\left[\frac{1}{250n\pi + i(4\pi^2n^2 - 2{,}500)} + \frac{1}{-250n\pi + i(4\pi^2n^2 - 2{,}500)}\right]$$

$$= -\frac{2E_0(4\pi^2n^2 - 2{,}500)}{(250n\pi)^2 + (4\pi^2n^2 - 2{,}500)^2} \qquad n \text{ odd}$$

$$b_n = i(D_n - D_{-n}) = E_0\left[\frac{1}{250n\pi + i(4\pi^2n^2 - 2{,}500)} - \frac{1}{-250n\pi + i(4\pi^2n^2 - 2{,}500)}\right]$$

$$= \frac{500\pi nE_0}{(250n\pi)^2 + (4\pi^2n^2 - 2{,}500)^2} \qquad n \text{ odd}$$

---

† Because of the presence of the condenser, the impedance for the DC component, or component of zero frequency, is infinite. Hence the term $\frac{1}{2}E_0$ in the expansion of $E(t)$ makes no contribution to the steady-state current.

EXERCISES

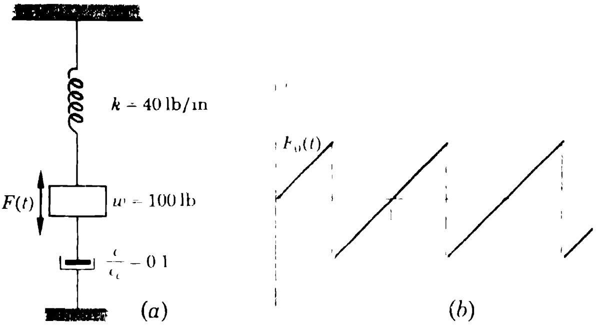
1 In Example 2, discuss the problem of determining the complete motion, transient as well as steady-state

2 In Example 3, why is the current $I(t)$ continuous when the impressed voltage $E(t)$ is discontinuous? Is the charge $Q(t)$ continuous?

3 In Example 2, determine the steady-state motion if the amount of friction is doubled and the spring is changed to one of modulus 120 lb/in

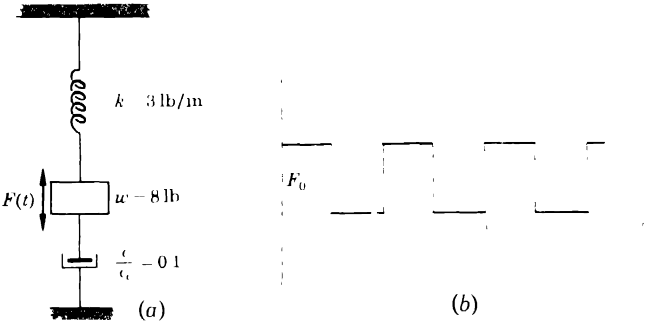
4 Determine the steady-state motion of the system shown in Fig. 6.10

FIGURE 6.10



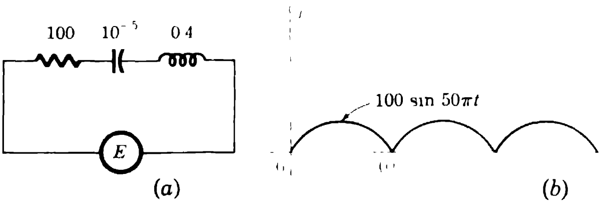
5 Determine the steady-state motion of the system shown in Fig. 6.11

FIGURE 6.11



6 Determine the steady-state current in the circuit shown in Fig. 6.12

FIGURE 6.12



7 Determine the steady-state current in the circuit shown in Fig 6 13

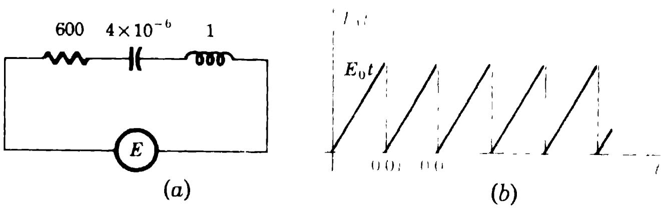


FIGURE 6 13

8 If $f(t) = A_1 \sin(\omega t - \delta_1) + A_2 \sin(2\omega t - \delta_2) + A_3 \sin(3\omega t - \delta_3) + \cdots$, show that

$$f(t)\Big|_{\text{rms}} = \sqrt{\tfrac{1}{2}\sum_{n=1}^{\infty} A_n^2}$$

9 If $f_1(t) = \sum_{n=-\infty}^{n=\infty} c_n e^{n i\pi t/p}$ and $f_2(t) = \sum_{n=-\infty}^{n=\infty} d_n e^{n i\pi t/p}$ are two functions of period $2p$, show that the average value of the product $f_1(t)f_2(t)$ over one period is $\sum_{n=-\infty}^{n=\infty} c_n d_{-n}$

10 If $f(t)$ is the periodic function whose definition in one period is

$$f(t) = |t| \qquad -\pi < t < \pi$$

find the solution of each of the following equations satisfying the indicated conditions

a $y'' - y = f(t)$ $\quad y_0 = y_0' = 0$ **b** $y'' + 3y' + 2y = f(t)$ $\quad y_0 = y_0' = 0$

c $y'' + 4y = f(t)$ $\quad y_0 = y_0' = 0$ **d** $y'' + 9y = f(t)$ $\quad y_0 = y_0' = 0$

6.6

Harmonic analysis

From time to time in applied work it is necessary to construct the Fourier expansion of a function defined by a table of values instead of by an analytic expression Various methods have been devised for doing this, comprising collectively the field of **harmonic analysis.*** Of these, the simplest and most obvious is the evaluation of the integrals in the Euler formulas by means of the trapezoidal rule Since this method is quite satisfactory for the occasional applications confronting the average worker and need be improved upon only for the person who must handle numerical functions regularly, it is the only method we shall discuss.

Because so many problems in harmonic analysis involve functions, such as meteorological or economic quantities, whose period is either a day or a year, it is customary to assume that data are available at intervals of $\frac{1}{12}$, $\frac{1}{24}$, or sometimes $\frac{1}{48}$ of a period. Accordingly, we shall consider a function $f(t)$ of period $2p$ for which values are available at intervals of

---

* For a more detailed discussion of harmonic analysis see, for instance, E T Whittaker and G Robinson, "The Calculus of Observations," pp 260–283, Blackie & Son, Ltd., Glasgow, 1937

$$\Delta t = \frac{2p}{24} = \frac{p}{12}$$

Now any function $f(t)$ can be expressed as the sum of an even function and an odd function simply by writing

$$f(t) = \frac{f(t) + f(-t)}{2} + \frac{f(t) - f(-t)}{2} = g(t) + h(t), \text{ say,}$$

since $g(t)$ is clearly even and $h(t)$ is clearly odd. Hence the cosine terms in the expansion of $f(t)$ are just the terms in the half-range cosine expansion of $g(t)$, and the sine terms in the expansion of $f(t)$ are just the terms in the half range sine expansion of $h(t)$. For the even function $g(t)$ we have, as usual,

$$a_n = \frac{2}{p}\int_0^p g(t)\cos\frac{n\pi t}{p}\,dt$$

or, applying the trapezoidal rule, with $\Delta t = p/12$,

$$a_n = \frac{2}{p}\left[\frac{p}{12}\left(\frac{g_0}{2}\cos\frac{n\pi}{p}0 + g_1\cos\frac{n\pi}{p}\frac{p}{12} + \cdots + g_{11}\cos\frac{n\pi}{p}\frac{11p}{12} + \frac{g_{12}}{2}\cos\frac{n\pi}{p}\frac{12p}{12}\right)\right]$$

$$(1)\qquad = \frac{1}{6}\left[\frac{g_0}{2} + g_1\cos\frac{n\pi}{12} + \cdots + g_{11}\cos\frac{11n\pi}{12} + \frac{g_{12}}{2}\cos n\pi\right]$$

The cosine factors in the last expression can be evaluated once and for all and combined with the other numerical factors, including the $\frac{1}{2}$ in the definition of $g(t)$, to yield a set of weights by which the successive values of the sum $f(t) + f(-t)$ are to be multiplied before they are added. The weights involved in the calculation of the first ten a's are given in Table 6.1.

Similarly, for $h(t)$ we have

$$b_n = \frac{2}{p}\int_0^p h(t)\sin\frac{n\pi t}{p}\,dt$$

$$= \frac{2}{p}\left[\frac{p}{12}\left(\frac{h_0}{2}\sin\frac{n\pi}{p}0 + h_1\sin\frac{n\pi}{p}\frac{p}{12} + \cdots + h_{11}\sin\frac{n\pi}{p}\frac{11p}{12} + \frac{h_{12}}{2}\sin\frac{n\pi}{p}\frac{12p}{12}\right)\right]$$

$$(2)\qquad = \frac{1}{6}\left[h_1\sin\frac{n\pi}{12} + \cdots + h_{11}\sin\frac{11n\pi}{12}\right]$$

The weights required for the evaluation of this expression are shown in Table 6.2.

table 6 1

Terms	Weights for the determination of a_n										
	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
$g_0† = f_0$	0.08333	0.08333	0.08333	0.08333	0.08333	0.08333	0.08333	0.08333	0.08333	0.08333	0.08333
$g_1 = f_1 + f_{-1}$	0.08333	0.08049	0.07217	0.05893	0.04167	0.02157	0.00000	−0.02157	−0.04167	−0.05893	−0.07217
$g_2 = f_2 + f_{-2}$	0.08333	0.07217	0.04167	0.00000	−0.04167	−0.07217	−0.08333	−0.07217	−0.04167	0.00000	0.04167
$g_3 = f_3 + f_{-3}$	0.08333	0.05893	0.00000	−0.05893	−0.08333	−0.05893	0.00000	0.05893	0.08333	0.05893	0.00000
$g_4 = f_4 + f_{-4}$	0.08333	0.04167	−0.04167	−0.08333	−0.04167	0.04167	0.08333	0.04167	−0.04167	−0.08333	−0.04167
$g_5 = f_5 + f_{-5}$	0.08333	0.02157	−0.07217	−0.05893	0.04167	0.08049	0.00000	−0.08049	−0.04167	0.05893	0.07217
$g_6 = f_6 + f_{-6}$	0.08333	0.00000	−0.08333	0.00000	0.08333	0.00000	−0.08333	0.00000	0.08333	0.00000	−0.08333
$g_7 = f_7 + f_{-7}$	0.08333	−0.02157	−0.07217	0.05893	0.04167	−0.08049	0.00000	0.08049	−0.04167	−0.05893	0.07217
$g_8 = f_8 + f_{-8}$	0.08333	−0.04167	−0.04167	0.08333	−0.04167	−0.04167	0.08333	−0.04167	−0.04167	0.08333	−0.04167
$g_9 = f_9 + f_{-9}$	0.08333	−0.05893	0.00000	0.05893	−0.08333	0.05893	0.00000	−0.05893	0.08333	−0.05893	0.00000
$g_{10} = f_{10} + f_{-10}$	0.08333	−0.07217	0.04167	0.00000	−0.04167	0.07217	−0.08333	0.07217	−0.04167	0.00000	0.04167
$g_{11} = f_{11} + f_{-11}$	0.08333	−0.08049	0.07217	−0.05893	0.04167	−0.02157	0.00000	0.02157	−0.04167	0.05893	−0.07217
$g_{12} = f_{12} + f_{-12}$	0.04167	−0.04167	0.04167	−0.04167	0.04167	−0.04167	0.04167	−0.04167	0.04167	−0.04167	0.04167

† With the exception of g_0, these entries are equal to twice the corresponding a's in Eq (1), since, for convenience the factor $\frac{1}{2}$ in the definition of $g(t)$ has been incorporated in the weights

table 6 2

Terms	Weights for the determination of b_n									
	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
h_1† $= f_1 - f_{-1}$	0 02157	0 04167	0 05893	0 07217	0 08049	0 08333	0 08049	0 07217	0 05893	0 04167
$h_2 = f_2 - f_{-2}$	0 04167	0 07217	0 08333	0 07217	0 04167	0 00000	−0 04167	−0 07217	−0 08333	−0 07217
$h_3 = f_3 - f_{-3}$	0 05893	0 08333	0 05893	0 00000	−0 05893	−0 08333	−0 05893	0 00000	0 05893	0 08333
$h_4 = f_4 - f_{-4}$	0 07217	0 07217	0 00000	−0 07217	−0 07217	0 00000	0 07217	0 07217	0 00000	−0 07217
$h_5 = f_5 - f_{-5}$	0 08049	0 04167	−0 05893	−0 07217	0 02157	0 08333	0 02157	−0 07217	−0 05893	0 04167
$h_6 = f_6 - f_{-6}$	0 08333	0 00000	−0 08333	0 00000	0 08333	0 00000	−0 08333	0 00000	0 08333	0 00000
$h_7 = f_7 - f_{-7}$	0 08049	−0 04167	−0 05893	0 07217	0 02157	−0 08333	0 02157	0 07217	−0 05893	−0 04167
$h_8 = f_8 - f_{-8}$	0 07217	−0 07217	0 00000	0 07217	−0 07217	0 00000	0 07217	−0 07217	0 00000	0 07217
$h_9 = f_9 - f_{-9}$	0 05893	−0 08333	0 05893	0 00000	−0 05893	0 08333	−0 05893	0 00000	0 05893	−0 08333
$h_{10} = f_{10} - f_{-10}$	0 04167	−0 07217	0 08333	−0 07217	0 04167	0 00000	−0 04167	0 07217	−0 08333	0 07217
$h_{11} = f_{11} - f_{-11}$	0 02157	−0 04167	0 05893	−0 07217	0 08049	−0 08333	0 08049	−0 07217	0 05893	−0 04167

† These entries are equal to twice the corresponding h's in Eq (2), since, for convenience, the factor ½ in the definition of $h(t)$ has been incorporated in the weights

EXAMPLE 1

Find a_1 for the periodic function whose definition in one period is

$$\begin{array}{lllll}
y_0 = -2.0000 & y_5 = -0.9236 & y_{10} = -0.1944 & y_{15} = 0.1875 & y_{20} = 0.2222 \\
y_1 = -1.7569 & y_6 = -0.7500 & y_{11} = -0.0903 & y_{16} = 0.2222 & y_{21} = 0.1875 \\
y_2 = -1.5278 & y_7 = -0.5903 & y_{12} = 0.0000 & y_{17} = 0.2431 & y_{22} = 0.1389 \\
y_3 = -1.3125 & y_8 = -0.4444 & y_{13} = 0.0764 & y_{18} = 0.2500 & y_{23} = 0.0764 \\
y_4 = -1.1111 & y_9 = -0.3125 & y_{14} = 0.1389 & y_{19} = 0.2431 & y_{24} = 0.0000
\end{array}$$

Interpreting the mid-ordinate y_{12} to be f_0 in our general discussion and using the weights in the column headed $n = 1$ in Table 6.1, we find without difficulty

Term	Weight	Product
$0.0000 = 0.0000$	0.08333	0.00000
$0.0764 - 0.0903 = -0.0139$	0.08049	−0.00112
$0.1389 - 0.1944 = -0.0555$	0.07217	−0.00401
$0.1875 - 0.3125 = -0.1250$	0.05893	−0.00737
$0.2222 - 0.4444 = -0.2222$	0.04167	−0.00926
$0.2431 - 0.5903 = -0.3472$	0.02157	−0.00749
$0.2500 - 0.7500 = -0.5000$	0.00000	0.00000
$0.2431 - 0.9236 = -0.6805$	−0.02157	0.01469
$0.2222 - 1.1111 = -0.8889$	−0.04167	0.03704
$0.1875 - 1.3125 = -1.1250$	−0.05893	0.06630
$0.1389 - 1.5278 = -1.3889$	−0.07217	0.10024
$0.0764 - 1.7569 = -1.6805$	−0.08049	0.13526
$0.0000 - 2.0000 = -2.0000$	−0.04167	0.08334
	$a_1 =$	0.40762

In this simple illustration, y is actually the function $t - t^2$, $-1 \leq t \leq 1$, whose expansion we obtained in Sec. 6.3. The exact value of a_1, as read from Eq. (4), Sec. 6.3, is $4/\pi^2 = 0.4053$, so our approximation is in error by about $\frac{1}{2}$ of 1 per cent.

EXERCISES

1. Determine by harmonic analysis the Fourier expansion of the circular arc $y = \sqrt{2\pi x - x^2}$ $(0 \leq x \leq 2\pi)$ through a_6 and b_6.
2. The following table gives the cylinder pressure in pounds per square inch for a certain 4-cycle internal combustion engine at 30° increments of the crank angle θ.

θ	P	θ	P	θ	P	θ	P	θ	P
0	200	150	45	300	0	450	0	600	5
30	350	180	20	330	0	480	0	630	11
60	167	210	6	360	0	510	0	660	30
90	102	240	0	390	0	540	0	690	90
120	65	270	0	420	0	570	0	720	200

Determine the harmonic analysis of the pressure through a_6 and b_6.

3. The normal maximum and minimum temperatures at New York City on the first and fifteenth of each month are given in the following table:

	Max	Min		Max	Min		Max	Min
Jan 1	38	26	May 1	63	48	Sept 1	77	64
15	37	24	15	68	52	15	74	60
Feb 1	37	24	June 1	73	57	Oct 1	69	55
15	38	24	15	77	60	15	64	49
Mar 1	41	26	July 1	80	64	Nov 1	57	43
15	45	30	15	82	66	15	51	37
Apr 1	51	36	Aug 1	82	67	Dec 1	45	32
15	57	42	15	80	67	15	41	29

Neglecting the slight irregularities in the spacing of the data, determine the harmonic analysis of the maximum temperature and of the minimum temperature.

4 By evaluating the complex form of the Fourier series of $f(t)$ at the points $t = 0, \pi/m, 2\pi/m, \ldots, (2m-1)\pi/m$ and using the fact that $e^{i\pi} = -1$ show that

$$f(0) - f\left(\frac{\pi}{m}\right) + f\left(\frac{2\pi}{m}\right) - \cdots - f\left(\frac{2m-1}{m}\pi\right)$$
$$= 2m(\cdots + c_{-5m} + c_{-3m} + c_{-m} + c_m + c_{3m} + c_{5m} + \cdots)$$
$$= 2m(a_m + a_{3m} + a_{5m} + \cdots)$$

Explain how this formula can be used to determine approximately the coefficients of the cosine terms in the Fourier expansion of $f(t)$.

5 By evaluating the complex form of the Fourier series of $f(t)$ at the points $t = \pi/2m, 3\pi/2m, 5\pi/2m, \ldots, (4m-1)\pi/2m$ and using the fact that $e^{i\pi/2} = i$, show that

$$f\left(\frac{\pi}{2m}\right) - f\left(\frac{3\pi}{2m}\right) + f\left(\frac{5\pi}{2m}\right) - \cdots - f\left(\frac{4m-1}{2m}\pi\right)$$
$$= 2mi(\cdots - c_{-5m} + c_{-3m} - c_{-m} + c_m - c_{3m} + c_{5m} - \cdots)$$
$$= 2m(b_m - b_{3m} + b_{5m} - \cdots)$$

Explain how this formula can be used to determine approximately the coefficients of the sine terms in the Fourier expansion of $f(t)$.

6.7

The Fourier integral as the limit of a Fourier series

The properties of Fourier series we have thus far developed are adequate to accomplish the expansion of any periodic function satisfying the Dirichlet conditions and, in connection with the theory of Chap. 5, enable us to find the response of numerous mechanical and electrical systems to general periodic disturbances. On the other hand, in many problems the impressed force or voltage is nonperiodic rather than periodic, a single unrepeated pulse, for instance. Functions of this sort cannot be handled directly through the use of Fourier series, for such series necessarily define only periodic functions. However, by investigating the limit (if any) which is approached by a Fourier series as the period of the given function becomes infinite, a suitable representation for nonperiodic functions can perhaps be obtained. An

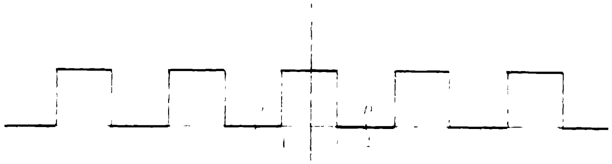


FIGURE 6.14
A periodic function of period $2p = 4$

example is probably the best way to introduce the theory of this procedure.

Consider, then, the function $f_p(t)$ shown in Fig. 6.14 in the limit as $p \to \infty$, as suggested by Fig. 6.15. This function is clearly even, and thus its Fourier expansion contains only cosine terms, i.e.,

(1)
$$f_p(t) = \frac{a_0}{2} + a_1 \cos\frac{\pi t}{p} + a_2 \cos\frac{2\pi t}{p} + \cdots + a_n \cos\frac{n\pi t}{p} + \cdots$$

where

(2a)
$$a_0 = \frac{2}{p}\int_0^1 1 \; dt = \frac{2}{p}$$

and

(2b)
$$a_n = \frac{2}{p}\int_0^1 1 \; \cos\frac{n\pi t}{p}\, dt = \frac{2}{p} \cdot \frac{\sin(n\pi t/p)}{n\pi/p}\bigg|_0^1 = \frac{2}{p}\,\frac{\sin(n\pi/p)}{n\pi/p}$$

It will help us now to understand what happens as $p \to \infty$ if

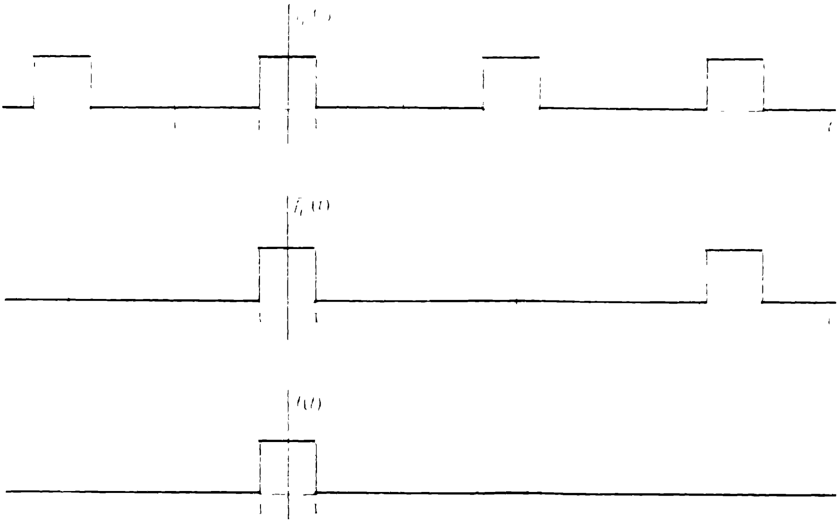


FIGURE 6.15
Plot illustrating the nonperiodic limit of a sequence of periodic functions whose periods become infinite.

we plot the "spectrum" of $f_p(t)$; that is, plot a_n as a function of the frequency

$$\omega_n = \frac{n\pi}{p} \text{ rad/unit time}$$

for different values of p. Introducing the symbol ω_n in Eq. (2b), we then have

(3) $$a_n = \frac{2}{p} \cdot \frac{\sin \omega_n}{\omega_n}$$

where successive values of n correspond to values of ω_n which differ by the constant amount

$$\Delta\omega = \frac{(n+1)\pi}{p} - \frac{n\pi}{p} = \frac{\pi}{p}$$

Hence the values of a_n are simply the ordinates of the curve

(4) $$y = \frac{2}{p} \frac{\sin \omega}{\omega} = \frac{2}{\pi} \frac{\sin \omega}{\omega} \Delta\omega$$

at successive ω-intervals of π/p, beginning at $\omega = 0$. These are shown in Fig. 6.16 for $p = 2$, 4, and 8.

As suggested by Fig. 6.16 and confirmed by Eq. (4), the

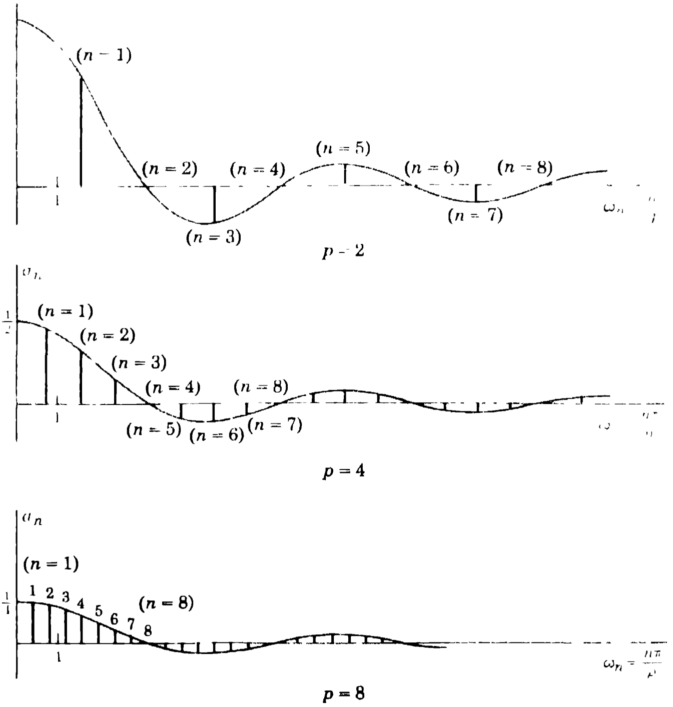


FIGURE 6.16
Plot illustrating the behavior of the Fourier coefficients of a function as the period of the function becomes infinite

coefficient plots, or spectra, for different values of p differ in only two respects

a In the vertical scale, which is inversely proportional to p (or directly proportional to $\Delta\omega$)

b In the horizontal interval between ordinates, which is also inversely proportional to p (or equal to $\Delta\omega$)

The fact that, as $p \to \infty$ (or $\Delta\omega \to 0$), the frequencies of the terms in (1) become more and more closely spaced and the coefficients approach zero suggests that this series, thought of as a function of p, is actually a sum of infinitesimals whose limit is an integral Indeed, this is true of Fourier series in general, as the following outline of steps reveals

If we begin with the complex exponential form of a Fourier series [Eqs (1) and (2), Sec 6 4]

$$f_p(t) = \sum_{n=-\infty}^{\infty} c_n e^{in\pi t/p}$$

$$c_n = \frac{1}{2p}\int_{-p}^{p} f_p(t)e^{-in\pi t/p}\,dt \equiv \frac{1}{2p}\int_{-p}^{p} f_p(s)e^{-in\pi s/p}\,ds$$

and substitute the second expression for c_n into $f_p(t)$, we obtain

$$f_p(t) = \sum_{n=-\infty}^{\infty}\left[\frac{1}{2p}\int_{-p}^{p} f_p(s)e^{-in\pi s/p}\,ds\right]e^{in\pi t/p}$$

$$= \sum_{n=-\infty}^{\infty}\left[\frac{1}{2\pi}\int_{-p}^{p} f_p(s)e^{-in\pi s/p}\,ds\right]e^{in\pi t/p}\,\frac{\pi}{p}$$

Now, as above, let us denote the frequency of the general term by

$$\omega_n = \frac{n\pi}{p}$$

and the difference in frequency between successive terms by

$$\Delta\omega = \frac{\pi}{p}$$

Then $f_p(t)$ can be written

(5) $$f_p(t) = \sum_{n=-\infty}^{\infty}\left[\frac{1}{2\pi}e^{i\omega_n t}\int_{-p}^{p} f_p(s)e^{-i\omega_n s}\,ds\right]\Delta\omega$$

If we now define

(6) $$F(\omega) = \frac{1}{2\pi}e^{i\omega t}\int_{-p}^{p} f_p(s)e^{-i\omega s}\,ds$$

Eq (5) becomes simply

(7) $$f_p(t) = \sum_{n=-\infty}^{\infty} F(\omega_n)\,\Delta\omega$$

where ω_n is a point (the left-hand end point, in fact) in the nth subinterval $\Delta\omega$ Under very general conditions, the limit of a sum of the form (7) as $\Delta\omega \to 0$ is the integral

$$\int_{-\infty}^{\infty} F(\omega)\, d\omega$$

Hence, since $p \to \infty$ implies $\Delta\omega \to 0$, it follows that there is good reason to believe that as $p \to \infty$ the nonperiodic limit of $f_p(t)$ can be written as the integral

$$f(t) = \int_{-\infty}^{\infty} \left[\frac{1}{2\pi} e^{i\omega t} \int_{-\infty}^{\infty} f(s) e^{-i\omega s}\, ds\right] d\omega \tag{8}$$

Though our derivation of it has been far from complete,* the last result is actually a valid representation of the nonperiodic limit function $f(t)$, provided that, in every finite interval, $f(t)$ satisfies the Dirichlet conditions and that the improper integral

$$\int_{-\infty}^{\infty} |f(t)|\, dt$$

exists Under these conditions, the so-called **Fourier integral** (8) gives the value of $f(t)$ at all points where $f(t)$ is continuous and gives the average of the right- and left-hand limits of $f(t)$ at all points where $f(t)$ is discontinuous

The Fourier integral can be written in various forms For instance, we can write Eq (8) as

$$f(t) = \int_{-\infty}^{\infty} g(\omega) e^{i\omega t}\, d\omega \tag{9}$$

where

$$g(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(s) e^{-i\omega s}\, ds \tag{10}$$

These two expressions, in which the symmetry between $f(t)$ and its **coefficient function** $g(\omega)$ is unmistakable, constitute what is known as a **Fourier transform pair.**† The coefficient function $g(\omega)$ is, of course, completely equivalent to $f(t)$, since when it is known, $f(t)$ is determined through Eq (9) In effect, we thus have two different representations of the function of our discussion $f(t)$ in the time domain and $g(\omega)$ in the frequency domain In passing, we note that elaborate tables of Fourier transform pairs have been prepared for engineering use ‡

---

* The situation is actually not so simple as we have made it appear, for from (6) it is clear that the structure of the function $F(\omega)$ depends on p as well as upon ω Hence, as p increases, the function we are evaluating changes, and the elementary theory of the definite integral is not strictly applicable Moreover, the fact that the summation extends over an infinite range makes additional investigation of the limiting process necessary The modifications required for a rigorous justification of our conclusions can be found in more advanced texts, such as R V Churchill, "Fourier Series and Boundary Value Problems," 2d ed , pp 113–117, McGraw-Hill Book Company, New York, 1963

† Sometimes it is more convenient to associate the factor $1/2\pi$ with the integral for $f(t)$ instead of with the integral for $g(\omega)$ It is also possible to achieve a still more symmetric form by associating the factor $1/\sqrt{2\pi}$ with each of the integrals

‡ G A Campbell and R M Foster, "Fourier Integrals for Practical Applications," D Van Nostrand Company, Inc , Princeton, N J , 1948

If we choose, we can, of course, move the factor $e^{i\omega t}$ into the integrand of the inner integral in (8), since it does not involve the variable s of that integration. This gives

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s)e^{-i\omega(s-t)}\,ds\,d\omega \tag{11}$$

In this, we can replace the exponential by its trigonometric equivalent, getting

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s)[\cos\omega(s-t) - i\sin\omega(s-t)]\,ds\,d\omega$$

If we break this up into two integrals, we get

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s)\cos\omega(s-t)\,ds\,d\omega - \frac{i}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s)\sin\omega(s-t)\,ds\,d\omega$$

Now $\sin\omega(s-t)$ is an odd function of ω. Hence the second integral vanishes because of the ω-integration from $-\infty$ to ∞. This could have been foreseen, of course, since by hypothesis $f(t)$ is purely real. Thus we obtain the real trigonometric representation

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s)\cos\omega(s-t)\,ds\,d\omega \tag{12a}$$

Since the integrand of (12a) is an even function of ω, we need perform the ω-integration only between 0 and ∞, provided we multiply the result by 2. This gives us the modified form

$$f(t) = \frac{1}{\pi}\int_{0}^{\infty}\int_{-\infty}^{\infty} f(s)\cos\omega(s-t)\,ds\,d\omega \tag{12b}$$

If $f(t)$ is either an odd function or an even function, further simplifications are possible. To see this, we first expand the factor $\cos\omega(s-t)$ in the integrand of (12a), getting

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s)\cos\omega s\cos\omega t\,ds\,d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s)\sin\omega s\sin\omega t\,ds\,d\omega$$

and then write the inner integrals as the sums of integrals over $(-\infty,0)$ and $(0,\infty)$. Then

$$\begin{aligned} f(t) = {} & \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{0} f(s)\cos\omega s\cos\omega t\,ds\,d\omega \\ & + \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{0}^{\infty} f(s)\cos\omega s\cos\omega t\,ds\,d\omega \\ & + \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{0} f(s)\sin\omega s\sin\omega t\,ds\,d\omega \\ & + \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{0}^{\infty} f(s)\sin\omega s\sin\omega t\,ds\,d\omega \end{aligned}$$

Next we make the substitution $s = -z,\ ds = -dz$ in the integrals from $-\infty$ to 0

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{\infty}^{0} f(-z)\cos(-\omega z)\cos\omega t\,(-dz)\,d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{0}^{\infty} f(s)\cos\omega s\cos\omega t\,ds\,d\omega$$
$$+\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{\infty}^{0} f(-z)\sin(-\omega z)\sin\omega t\,(-dz)\,d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{0}^{\infty} f(s)\sin\omega s\sin\omega t\,ds\,d\omega \tag{13}$$

Now, if $f(t)$ is an even function, so that $f(-z) = f(z)$, the first integral in (13) becomes identical with the second when the minus sign attached to dz is used to reverse the order of the limits on the inner integral. Similarly, the third and fourth integrals turn out to be negatives of each other. Hence we have simply

$$f(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\int_{0}^{\infty} f(s)\cos\omega s\cos\omega t\,ds\,d\omega \qquad f(t)\text{ even} \tag{14}$$

This is called the **Fourier cosine integral** and is analogous to the half-range cosine expansion of a periodic function which is even.

If $f(t)$ is an odd function, so that $f(-z) = -f(z)$, then the first and second integrals in (13) cancel each other, and the third and fourth combine, giving

$$f(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\int_{0}^{\infty} f(s)\sin\omega s\sin\omega t\,ds\,d\omega \qquad f(t)\text{ odd} \tag{15}$$

This is the **Fourier sine integral** and is the analogue of the half-range sine expansion of an odd periodic function.

For some purposes it is convenient to have the Fourier cosine and sine integral representations displayed as transform pairs. Thus we can write (14) in the form

$$\left.\begin{aligned} f(t) &= \int_{-\infty}^{\infty} g(\omega)\cos\omega t\,d\omega \\ g(\omega) &= \frac{1}{\pi}\int_{0}^{\infty} f(s)\cos\omega s\,ds \end{aligned}\right\} \qquad f(t)\text{ even} \tag{14a}$$

and (15) in the form

$$\left.\begin{aligned} f(t) &= \int_{-\infty}^{\infty} g(\omega)\sin\omega t\,d\omega \\ g(\omega) &= \frac{1}{\pi}\int_{0}^{\infty} f(s)\sin\omega s\,ds \end{aligned}\right\} \qquad f(t)\text{ odd} \tag{15a}$$

Equations (14), (14*a*), (15), and (15*a*) can, of course, all be modified by performing the ω-integrations only from 0 to ∞ and multiplying the results by 2.

To illustrate the Fourier integral representation of a nonperiodic function, let us return to the isolated pulse which we considered briefly at the beginning of this section (Fig. 6.15)

Since this function is clearly even, we can use (14), getting

$$f(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\int_0^1 1 \cos \omega s \cos \omega t \, ds \, d\omega = \frac{1}{\pi}\int_{-\infty}^{\infty} \cos \omega t \left[\frac{\sin \omega s}{\omega}\right]_0^1 d\omega$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\cos \omega t \sin \omega}{\omega} \, d\omega$$

$$(16) \qquad = \frac{2}{\pi}\int_0^{\infty} \frac{\cos \omega t \sin \omega}{\omega} \, d\omega$$

where the last step follows because the integrand is an even function of ω Thus, although it is impossible to find an elementary antiderivative for the last integral, we know that, as a definite integral, it must equal 1 if t is between -1 and $+1$, must equal $\frac{1}{2}$ if $t = \pm 1$, and must vanish if t is numerically greater than 1

In the case of the Fourier-series representation of a periodic function it was a matter of some interest to determine how well the first few terms of the expansion represented the function (Fig 6 3) The corresponding problem in the nonperiodic case is to investigate how well the Fourier integral represents a function when only the components in the lower part of the (continuous) frequency range are taken into account Suppose, therefore, that we consider only the frequencies below ω_0 In this case, from (16) we have, as an approximation to $f(t)$, the finite integral

$$\frac{2}{\pi}\int_0^{\omega_0} \frac{\cos \omega t \sin \omega}{\omega} \, d\omega$$

Now $\qquad \cos a \sin b = \dfrac{\sin (a + b) - \sin (a - b)}{2}$

and thus we can write the last integral as

$$\frac{1}{\pi}\int_0^{\omega_0} \frac{\sin \omega(t + 1)}{\omega} \, d\omega - \frac{1}{\pi}\int_0^{\omega_0} \frac{\sin \omega(t - 1)}{\omega} \, d\omega$$

In the first of these terms let $\omega(t + 1) = u$, and in the second let $\omega(t - 1) = u$ Then, for our approximation to $f(t)$, we have

$$\frac{1}{\pi}\int_0^{\omega_0(t+1)} \frac{\sin u}{u} \, du - \frac{1}{\pi}\int_0^{\omega_0(t-1)} \frac{\sin u}{u} \, du$$

Although integrals of this form cannot be expressed in terms of elementary functions, they occur often enough in applied mathematics to have been named and tabulated Specifically,

$$\operatorname{Si}(x) \equiv \int_0^x \frac{\sin u}{u} \, du$$

is known as the **sine integral** function of x and is tabulated in numerous handbooks * Using this notation, the approximation to $f(t)$ can be written

$$\frac{1}{\pi} \operatorname{Si} \omega_0(t + 1) - \frac{1}{\pi} \operatorname{Si} \omega_0(t - 1)$$

---

* See, for instance, E Jahnke, F Emde, and F Losch, "Tables of Higher Functions," 6th ed , McGraw-Hill Book Company, New York, 1960

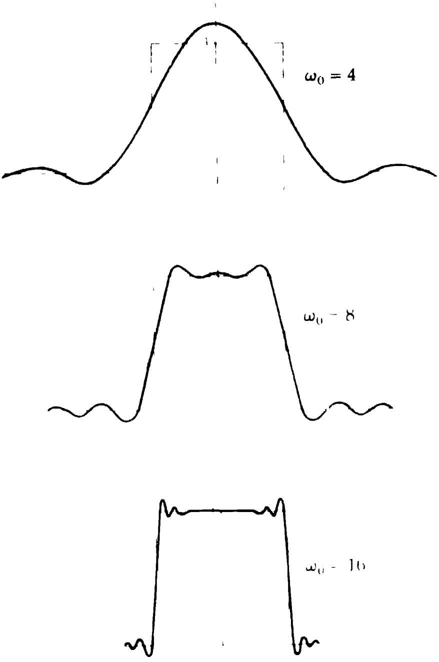


FIGURE 6.17
Plot showing the approximation of a function by its Fourier integral taken only over frequencies less than ω_0

Figure 6.17 shows this approximation for $\omega_0 = 4$, 8, and 16 rad/unit time. Physically speaking, these curves describe the output of an ideal low-pass filter, cutting off all frequencies above ω_0, when the input signal is an isolated rectangular pulse.

The Fourier integral representation of a nonperiodic function can be used in essentially the same way as the Fourier series representation of a periodic function in applications like those we considered in Sec. 6.5. For instance, if an electrical circuit is acted upon by a nonperiodic voltage $f(t)$ whose Fourier integral representation is [Eqs. (9) and (10)]

$$f(t) = \int_{-\infty}^{\infty} g(\omega)e^{i\omega t}\,d\omega \qquad \text{where} \qquad g(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(s)e^{-i\omega s}\,ds$$

we can still, for purposes of analysis, think of $f(t)$ as being the sum of an infinite number of complex voltages $e^{i\omega t}$. In fact, the only distinction between the periodic and nonperiodic cases is that in the latter the "spectrum" of $f(t)$ contains terms of *all* frequencies and the amplitude, or intensity, of the component of any given frequency ω is infinitesimal, namely,

$$g(\omega)\,d\omega$$

Now in Sec. 5.4 we saw that the current produced in a circuit of impedance $Z(\omega)$ by a complex voltage $E_0e^{i\omega t}$ is simply

$$\frac{E_0e^{i\omega t}}{Z(\omega)}$$

Hence, to find the current produced by the nonperiodic voltage $f(t)$, we need only divide the infinitesimal voltage

$$[g(\omega)\,d\omega]e^{i\omega t}$$

corresponding to the general frequency ω by the value of the impedance $Z(\omega)$ at that frequency and then "add," i e , integrate, all the infinitesimal currents thus obtained The result is simply

$$I(t) = \int_{-\infty}^{\infty} \frac{g(\omega)}{Z(\omega)} e^{i\omega t}\,d\omega \qquad \text{where} \qquad g(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(s)e^{-i\omega s}\,ds$$

and $Z(\omega)$ is the impedance of the circuit

A similar discussion of mechanical systems acted upon by nonperiodic forces, using the magnification ratio and phase shift [Eqs (10) and (11), Sec 5 3] instead of the complex impedance, can be based on Eq (12*a*)

EXERCISES

1 Make an amplitude-frequency plot for $p = 2$, 4, and 8 for the periodic function whose definition in one period is

a $f(t) = \begin{cases} 0 & -p < t < -1 \\ -1 & -1 < t < 0 \\ 1 & 0 < t < 1 \\ 0 & 1 < t < p \end{cases}$ **b** $f(t) = \begin{cases} 0 & -p < t < -1 \\ 1 + t & -1 < t < 0 \\ 1 - t & 0 < t < 1 \\ 0 & 1 < t < p \end{cases}$

c $f(t) = \begin{cases} 0 & -p < t < -1 \\ 1 - t^2 & -1 < t < 1 \\ 0 & 1 < t < p \end{cases}$ **d** $f(t) = \begin{cases} 0 & -p < t < -1 \\ \sin \pi t & -1 < t < 1 \\ 0 & 1 < t < p \end{cases}$

2 Make an amplitude-frequency plot for $p = 2$, 4, and 8 for the periodic function whose definition in one period is

a $f(t) = \begin{cases} e^{t} & -p < t < 0 \\ e^{-t} & 0 < t < p \end{cases}$ **b** $f(t) = \begin{cases} -e^{t} & -p < t < 0 \\ e^{-t} & 0 < t < p \end{cases}$

3 Find the Fourier integral representation of each of the following functions

a $f(t) = \begin{cases} e^{at} & t < 0 \\ e^{-at} & t > 0 \end{cases}$ **b** $f(t) = \begin{cases} 0 & t < 0 \\ e^{-at} & t > 0 \end{cases}$

c $f(t) = \begin{cases} \sin t & t^2 < \pi^2 \\ 0 & t^2 > \pi^2 \end{cases}$ **d** $f(t) = \begin{cases} \cos t & t^2 < \pi^2/4 \\ 0 & t^2 > \pi^2/4 \end{cases}$

e $f(t) = \begin{cases} 0 & -\infty < t < 0 \\ 1 & 0 < t < 1 \\ 0 & 1 < t < \infty \end{cases}$ **f** $f(t) = \begin{cases} 1 - t^2 & t^2 < 1 \\ 0 & t^2 > 1 \end{cases}$

4 Find the Fourier integral representation of the function

$$f(t) = \begin{cases} 0 & -\infty < t < -1 \\ -1 & -1 < t < 0 \\ 1 & 0 < t < 1 \\ 0 & 1 < t < \infty \end{cases}$$

and express the integral which approximates this function for frequencies between 0 and ω_0 in terms of sine integral functions

5 Find the Fourier integral representation of the function

$$f(t) = \begin{cases} 0 & -\infty < t < -1 \\ 1 + t & -1 < t < 0 \\ 1 - t & 0 < t < 1 \\ 0 & 1 < t < \infty \end{cases}$$

and express the integral which approximates this function for frequencies between 0 and ω_0 in terms of sine integral functions

6 Find the Fourier integral representation of the function

$$f(t) = \begin{cases} t & t^2 < 1 \\ 0 & t^2 > 1 \end{cases}$$

7 Show that $\displaystyle \frac{2}{\pi}\int_0^\infty \frac{(2 - \omega^2)\cos \omega t + 3\omega \sin \omega t}{(2 - \omega^2)^2 + 9\omega^2}\frac{\sin \omega}{\omega}\,d\omega$ is a particular integral of the equation $y'' + 3y' + 2y = f(t)$

where $$f(t) = \begin{cases} 1 & t^2 < 1 \\ 0 & t^2 > 1 \end{cases}$$

8 Find a particular integral of the equation $y'' + ay' + by = f(t)$

where $$f(t) = \begin{cases} 1 & t^2 < 1 \\ & t^2 > 1 \end{cases}$$

9 Find a particular integral of the equation $y'' + ay' + by = f(t)$ where $f(t)$ is the function described in Exercise 5

10 Find a particular integral of the equation $y'' + ay' + by = f(t)$ where $f(t)$ is the function described in Exercise 6

11 **a** Using the Fourier integral representation (12a) and the concepts of magnification ratio and phase angle, obtain a formula for the response of a mechanical system of one degree of freedom to a nonperiodic driving force

b Show that the response of an electrical circuit of impedance $Z(\omega)$ to a nonperiodic voltage

$$f(t) = \int_{-\infty}^{\infty} g(\omega)e^{i\omega t}\,d\omega$$

can be written

$$I(t) = 2\int_0^\infty \left[\mathcal{R}\left\{\frac{g(\omega)}{Z(\omega)}\right\}\cos \omega t + \mathcal{I}\left\{\frac{g(\omega)}{Z(\omega)}\right\}\sin \omega t \right] d\omega$$

where $\mathcal{R}\{g(\omega)/Z(\omega)\}$ and $\mathcal{I}\{g(\omega)/Z(\omega)\}$ denote, respectively, the real and the imaginary part of the complex expression $g(\omega)/Z(\omega)$

12 If we call $g(\omega)$ in Eq (10) the **Fourier transform** of $f(t)$, show that, if the various transforms exist, then

a The Fourier transform of $e^{\pm i\omega_0 t}f(t)$ is $g(\omega \mp \omega_0)$

b The Fourier transform of $f'(t)$ is $i\omega g(\omega)$

c The Fourier transform of $\displaystyle\int_{-\infty}^{t} f(t)\,dt$ is $g(\omega)/i\omega$

d The Fourier transform of $f_1(t)f_2(t)$ is

$$\int_{-\infty}^{\infty} g_1(u)g_2(\omega - u)\,du = \int_{-\infty}^{\infty} g_1(\omega - v)g_2(v)\,dv$$

where $g_1(\omega)$ and $g_2(\omega)$ are, respectively, the Fourier transforms of $f_1(t)$ and $f_2(t)$

13 Let $f(t)$ be a function which is identically zero outside the interval $(-1,1)$, so that the

Fourier transform of $f(t)$ is

$$T(f) = \frac{1}{2\pi}\int_{-1}^{1} f(t)e^{-i\omega t}\,dt$$

By repeated differentiation of $T(1)$, show that

$$T(t^n) = \frac{(i)^n}{\pi}\,\frac{d^nS(\omega)}{d\omega^n}$$

where $S(\omega) = (\sin\omega)/\omega$. Explain how this result can be used to obtain the Fourier transform of a single pulse defined between -1 and 1 by a convergent power series.*

14 Using the definitions of Exercise 13, show that

$$T(e^{n i\pi t}) = \frac{1}{\pi}\,S(\omega - n\pi)$$

Explain how this result can be used to obtain the Fourier transform of a single pulse defined between -1 and 1 by a Fourier series in either complex exponential or real trigonometric form.

15 If $f(t)$ is a pulse defined between -1 and 1 by either of the equivalent series

$$\sum_{n=0}^{\infty}(a_n\cos n\pi t + b_n\sin n\pi t) \qquad \sum_{n=-\infty}^{\infty} c_n e^{n i\pi t}$$

use the results of Exercise 14 to show that

$$\begin{aligned} a_n &= \pi[\phi(n\pi) + \phi(-n\pi)] \\ b_n &= i\pi[\phi(n\pi) - \phi(-n\pi)] \\ c_n &= \pi\phi(n\pi) \end{aligned}$$

where $\phi(\omega)$ is the Fourier transform of the pulse.

6.8

From the Fourier integral to the Laplace transform

In many applications of the Fourier integral the function to be represented is identically zero before some instant, usually $t = 0$. When this is the case, the general Fourier transform pair, given by Eqs. (9) and (10), Sec. 6.7, becomes the **unilateral Fourier transform pair**†

$$(1)\qquad f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(\omega)e^{i\omega t}\,d\omega \qquad g(\omega) = \int_{0}^{\infty} f(s)e^{-i\omega s}\,ds$$

Useful as this is in many applications, it is still inadequate to

---

* In particular problems, the book "Tables of the Function $\frac{\sin u}{u}$ and of Its First Eleven Derivatives," Harvard University Press, Cambridge, Mass., will be of considerable help.

† Note that, for later convenience, we have chosen to incorporate the factor $1/2\pi$ in the integral for $f(t)$ rather than in the integral for $g(\omega)$, as we did earlier.

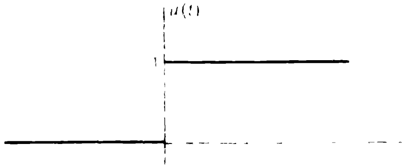


FIGURE 6.18
The unit step function $u(t)$

represent such a simple function as the so-called **unit step function** $u(t)$ (Fig 6 18)

$$u(t) = \begin{cases} 0 & t < 0 \\ 1 & t > 0 \end{cases}$$

In fact, for this function

$$g(\omega) = \int_0^\infty 1 \quad e^{-i\omega s}\, ds = \left.\frac{e^{-i\omega s}}{-i\omega}\right|_0^\infty = \left.\frac{\cos \omega s - i \sin \omega s}{-i\omega}\right|_0^\infty$$

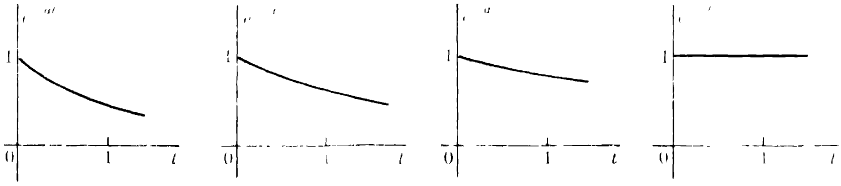
and this is completely meaningless, since both the cosine and sine oscillate without limit as their arguments become infinite

As an artifice to handle this case and others like it, the function e^{-at} is sometimes inserted in place of the unit step function. Now as we shall soon see, e^{-at} has a unilateral Fourier transform when a is positive Moreover, when a approaches zero, e^{-at}, considered for $t > 0$, approaches the unit step function (Fig 6 19) Hence, it is natural to hope that the order of the operations of letting a approach zero and taking the Fourier transform can be interchanged. *If* this is the case, then we can postpone letting $a \to 0$ until *after* the transform has been taken, and all will be well

In the present problem the development proceeds as follows Instead of transforming $u(t)$ we transform e^{-at}, getting

$$g(\omega) = \int_0^\infty e^{-as}e^{-i\omega s}\, ds = \left.\frac{e^{-(a+i\omega)s}}{-(a+i\omega)}\right|_0^\infty = \frac{1}{a+i\omega}$$

since the factor e^{-as} now ensures that the antiderivative vanishes



$a = \frac{3}{4}$ $a = \frac{1}{2}$ $a = \frac{1}{4}$ $a = 0$

FIGURE 6.19
Plot showing how e^{-at} $(a,t > 0)$ approaches the unit step function when a approaches zero

at the upper limit. Thus,

$$\begin{aligned}
f(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} g(\omega)e^{i\omega t}\,d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{i\omega t}}{a+i\omega}\,d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\cos\omega t + i\sin\omega t}{a+i\omega}\,\frac{a-i\omega}{a-i\omega}\,d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{(a\cos\omega t + \omega\sin\omega t) + i(a\sin\omega t - \omega\cos\omega t)}{a^2+\omega^2}\,d\omega
\end{aligned}$$

Now, the imaginary part of the integrand, namely,

$$\frac{a\sin\omega t - \omega\cos\omega t}{a^2+\omega^2}$$

is an odd function of ω and, hence, will vanish when integrated between the limits $-\infty$ and ∞. On the other hand, the real part of the integrand is an even function of ω, and thus we can write

$$f(t) = \frac{1}{\pi}\int_0^{\infty} \frac{a\cos\omega t + \omega\sin\omega t}{a^2+\omega^2}\,d\omega = \frac{1}{\pi}\int_0^{\infty} \frac{a\cos\omega t}{a^2+\omega^2}\,d\omega + \frac{1}{\pi}\int_0^{\infty} \frac{\omega\sin\omega t}{a^2+\omega^2}\,d\omega$$

In the first integral in the right member, let $\omega = az$. Then

$$f(t) = \frac{1}{\pi}\int_0^{\infty} \frac{\cos atz}{1+z^2}\,dz + \frac{1}{\pi}\int_0^{\infty} \frac{\omega\sin\omega t}{a^2+\omega^2}\,d\omega$$

We are now in a position to let a approach zero. As this happens,

$$f(t) \equiv e^{-at} \rightarrow u(t)$$

and thus we obtain

$$u(t) = \frac{1}{\pi}\int_0^{\infty} \frac{dz}{1+z^2} + \frac{1}{\pi}\int_0^{\infty} \frac{\sin\omega t}{\omega}\,d\omega = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \frac{\sin\omega t}{\omega}\,d\omega$$

This establishes the value of another definite integral, without benefit of an antiderivative.

The use we have just made of the so-called **convergence factor** e^{-at} is both artificial and clumsy, and it would be desirable to make this procedure more systematic. To do this, let us define

$$F(t) = \begin{cases} 0 & t < 0 \\ e^{-at}f(t) & t > 0 \end{cases}$$

where $f(t)$ is the function of actual interest. Then, applying the unilateral Fourier transformation to $F(t)$, which surely satisfies the necessary conditions if $f(t)$ does, we have, for $t > 0$,

$$F(t) = e^{-at}f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(\omega)e^{i\omega t}\,d\omega$$

where
$$\begin{aligned}
g(\omega) &= \int_0^{\infty} F(s)e^{-i\omega s}\,ds = \int_0^{\infty} e^{-as}f(s)e^{-i\omega s}\,ds \\
&= \int_0^{\infty} f(s)e^{-(a+i\omega)s}\,ds
\end{aligned}$$

We can now multiply both sides of the expression for $F(t)$ by e^{at}, getting

$$f(t) = \frac{e^{at}}{2\pi}\int_{-\infty}^{\infty} g(\omega)e^{i\omega t}\,d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(\omega)e^{(a+i\omega)t}\,d\omega$$

Moreover, from the last form of the expression for $g(\omega)$ it is clear that ω enters the analysis only through the binomial $a + i\omega$. To emphasize this fact, we shall write $g(a + i\omega)$ instead of $g(\omega)$ Then the equations of the transform pair become

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(a + i\omega)e^{(a+i\omega)t}\,d\omega$$

$$g(a + i\omega) = \int_0^{\infty} f(s)e^{-(a+i\omega)s}\,ds$$

Finally, let us put $a + i\omega = \sigma$, noting that

$$d\omega = \frac{d(a + i\omega)}{i} = \frac{d\sigma}{i}$$

and that when $\omega = -\infty$, $\sigma = a - i\infty$ and when $\omega = \infty$, $\sigma = a + i\infty$ Then we have the pair of equations

$$f(t) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} g(\sigma)e^{\sigma t}\,d\sigma \qquad g(\sigma) = \int_0^{\infty} f(s)e^{-\sigma s}\,ds\dagger$$

These constitute a **Laplace transform pair.**‡ The function $g(\sigma)$ is known as the **Laplace transform** of $f(t)$ The integral for $f(t)$ is known as the **complex inversion integral.**

We have thus naturally and inevitably encountered the Laplace transformation through our attempt to provide the unilateral Fourier transformation with a "built-in" convergence factor This transformation is the foundation of the modern form of the **operational calculus,** which was originated in quite another form by the English electrical engineer Oliver Heaviside around 1890 In the next chapter we shall develop an extensive list of formulas for the use of the Laplace transform itself, although the meaning and use of the inversion integral we must leave to the chapters on complex-variable theory

---

† Most writers use t rather than s as the dummy variable in this integral, and use s rather than σ in both integrals In the next chapter we shall follow this convention

‡ Named for Pierre Simon de Laplace (1749-1827), who used such transforms in his researches in the theory of probability

CHAPTER SEVEN

The Laplace Transformation

7.1

Theoretical preliminaries

In the last chapter we traced the evolution of the Laplace transformation from the unilateral Fourier integral. Our development made it clear that, for the Laplace transform of $f(t)$ to exist and for $f(t)$ to be recoverable from its transform, it is sufficient that

a In every interval of the form $0 \leqq t_1 \leqq t \leqq t_2$, $f(t)$ be bounded and have at most a finite number of maxima and minima and a finite number of finite discontinuities.

b There exist a real constant a such that the improper integral

$$\int_0^\infty |e^{-at}f(t)|\,dt = \int_0^\infty e^{-at}|f(t)|\,dt \text{ is convergent}$$

Functions satisfying condition **a** we shall henceforth describe as **piecewise regular.**

Condition **b** is frequently replaced by the stronger, i.e., more restrictive, condition that

b′ There exist constants α, M, and T such that

$$e^{-\alpha t}|f(t)| < M \qquad \text{for all } t > T$$

Functions which satisfy condition **b′** are usually described as being of **exponential order.**

Obviously, if $e^{-\alpha t}|f(t)| < M$, then $e^{-\alpha_1 t}|f(t)| < M$ for all $\alpha_1 > \alpha$. Thus the α required by condition **b′** is not unique. The greatest lower bound α_0 of the set of all α's which can be used in condition **b′** is often called the **abscissa of convergence** of $f(t)$. Under this definition, it is evident that the abscissa of convergence α_0 may not itself be one of the α's which will serve in condition **b′**. For instance, if $f(t) \equiv t$, then, for every positive α and no others,

$$e^{-\alpha t}|f(t)| \equiv te^{-\alpha t}$$

remains bounded and in fact approaches zero as t becomes infinite. Obviously the greatest lower bound of the set of all positive numbers is the number 0. Hence, in this case $\alpha_0 = 0$, even though for α_0 itself

$$e^{-\alpha_0 t}|f(t)| \equiv t$$

increases beyond all bounds as $t \to \infty$. In passing, we note that the abscissa of convergence of a function may be negative. For example, for $f(t) \equiv e^{-2t}$ the abscissa of convergence is -2, since as $t \to \infty$

$$e^{-\alpha t}|f(t)| \equiv e^{-\alpha t}e^{-2t}$$

is bounded for all values of α equal to or greater than -2 but for none less than -2.

Since $e^{-\alpha t}|f(t)| < M$ implies that $|f(t)| < Me^{\alpha t}$, it is clear that, if a function is of exponential order, its absolute value need not remain bounded as $t \to \infty$, but it must not increase more rapidly than some constant multiple of a simple exponential function of t. As the particular function $f(t) \equiv \sin e^{t^2}$ shows, *the derivative of a function of exponential order is not necessarily of exponential order.* On the other hand, it is not difficult to show that *if $f(t)$ is piecewise regular and of exponential order, then $\int_0^t f(t)\,dt$ is also of exponential order.*

With a function $f(t)$ satisfying either conditions **a** and **b** or conditions **a** and **b′**, the Laplace transformation associates a function of s, which we shall denote by $\mathcal{L}\{f(t)\}$† or simply by $\mathcal{L}\{f\}$. This is defined by the formula

$$(1) \qquad \mathcal{L}\{f(t)\} = \int_0^\infty f(t)e^{-st}\,dt‡$$

The function $f(t)$ whose Laplace transform is a given function of s, say $\phi(s)$, we shall call the **inverse** of $\phi(s)$ and shall denote by the symbol $\mathcal{L}^{-1}\{\phi(s)\}$. From the concluding discussion of the last chapter we have good reason to believe that the function having $\phi(s)$ for its transform is given by the complex inversion integral

$$(2) \qquad f(t) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \phi(s)e^{st}\,ds$$

where s is the complex variable $a + i\omega$, but we shall make no use of this fact in the present chapter. Indeed, in this chapter we shall regard s as a real-valued parameter.

It is obvious that the derivation of the fundamental properties of the Laplace transform will involve manipulation of the definitive integral (1) This integral is clearly improper, since its upper limit is infinite, and it may also be improper because of discontinuities of $f(t)$ at one or more points in the range of integration However, inasmuch as $f(t)$ is assumed to be piecewise regular, these discontinuities can be at worst finite jumps which can easily be handled by breaking up the range of integration into subranges whose end points are the points of discontinuity. We shall, therefore, usually not pay explicit attention to the possible jumps of $f(t)$ Questions associated with the infinite upper limit in (1) are more serious, however, and cannot be passed over so lightly.

---

† Many writers consistently use only small letters to denote functions of t and use the corresponding capital letters to denote the transforms of these functions. Thus what we shall write as $\mathcal{L}\{f(t)\}$ is often written as $F(s)$.

‡ Clearly, the variable of integration t is a *dummy variable* and can be replaced at pleasure by any other symbol. From time to time we shall find it convenient to do this in our work.

At the outset, we recall that by an integral of the form

$$\int_a^\infty h(s,t)\,dt \tag{3}$$

we mean $\displaystyle\lim_{b\to\infty}\int_a^b h(s,t)\,dt$

and that for this limit to exist for a particular value of s, say $s = s_1$, it must be possible to show that for any $\epsilon > 0$ there exists a number B such that

$$\left|\int_a^\infty h(s_1,t)\,dt - \int_a^b h(s_1,t)\,dt\right| = \left|\int_b^\infty h(s_1,t)\,dt\right| < \epsilon$$

for all values of $b > B$ The number B will, of course, depend on ϵ and in general will also depend on s_1, the particular value of s under consideration It may happen, however, that one and the same number B will serve uniformly, or equally well, for all members of some set of s-values If and only if this is the case, the integral (3) is said to converge uniformly, or to have the property of **uniform convergence,** over that particular set of s-values

The importance of uniform convergence is apparent from the following theorems, which we shall have to use in this chapter but whose proofs we leave to more advanced texts *

THEOREM 1

If $g(s,t)$ is a continuous function of s and t for $\alpha \leqq s \leqq \beta$ and $t \geqq a$, if $f(t)$ is at least piecewise regular for $t \geqq a$, and if the integral $G(s) = \int_a^\infty f(t)g(s,t)\,dt$ converges uniformly over the interval $\alpha \leqq s \leqq \beta$, then $G(s)$ is a continuous function of s for $\alpha \leqq s \leqq \beta$

Since the definitive property of a continuous function is that

$$\lim_{s\to s_0} G(s) = G(s_0)$$

this theorem states, in effect, that under the appropriate conditions the limit of $G(s)$ can be found by taking the limit inside the integral sign

---

* See, for instance, H S Carslaw, "Fourier Series," pp 198–201, Dover Publications, Inc , New York, 1930

THEOREM 2

If $g(s,t)$ is a continuous function of s and t for $\alpha \leqq s \leqq \beta$ and $t \geqq a$, if $f(t)$ is at least piecewise regular for $t \geqq a$, and if the integral $G(s) = \int_a^\infty f(t)g(s,t)\,dt$ converges uniformly over the interval $\alpha \leqq s \leqq \beta$, then

$$\int_\alpha^\beta G(s)\,ds \equiv \int_\alpha^\beta \int_a^\infty f(t)g(s,t)\,dt\,ds = \int_a^\infty \int_\alpha^\beta f(t)g(s,t)\,ds\,dt$$

In words, this theorem states that under the appropriate conditions the integral of $G(s)$ can be found by integrating inside the integral sign

THEOREM 3

If $g(s,t)$ and $g_s(s,t) \equiv \dfrac{\partial g(s,t)}{\partial s}$ are continuous functions of s and t for $\alpha \leqq s \leqq \beta$ and $t \geqq a$, if $f(t)$ is at least piecewise regular for $t \geqq a$, if the integral

$$G(s) = \int_a^\infty f(t)g(s,t)\,dt$$

converges, and if $\int_a^\infty f(t)g_s(s,t)\,dt$ converges uniformly over the interval $\alpha \leqq s \leqq \beta$, then

$$G'(s) \equiv \frac{d}{ds}\int_a^\infty f(t)g(s,t)\,dt = \int_a^\infty f(t)g_s(s,t)\,dt$$

for all values of s such that $\alpha \leqq s \leqq \beta$

In words, Theorem 3 states that, under the appropriate conditions, the derivative of $G(s)$ can be found by differentiating inside the integral sign

Obviously, if we take $g(s,t)$ to be the continuous function e^{-st} and take $a = 0$, the integral $G(s)$ referred to in the last three theorems is precisely the Laplace transform of the function $f(t)$ However, before we can apply these theorems to our work we must determine under what conditions the Laplace transform integral converges uniformly We begin by proving the following weaker result

THEOREM 4

If $f(t)$ is piecewise regular and of exponential order, then

$$\mathcal{L}\{f(t)\} = \int_0^\infty f(t)e^{-st}\,dt$$

converges absolutely for any value of s greater than the abscissa of convergence α_0 of $f(t)$

PROOF To establish this theorem, we must show that

$$(4) \qquad \lim_{b\to\infty}\int_0^b |f(t)e^{-st}|\,dt = \lim_{b\to\infty}\int_0^b |f(t)|e^{-st}\,dt$$

exists, and to do this it is necessary that we have an upper bound for $|f(t)|$ for $t \geqq 0$ Now, by hypothesis, $f(t)$ is of exponential order and, therefore, has an abscissa of convergence α_0. Hence, there exist numbers M_1 and T such that for all $t > T$ and any α greater than but bounded from α_0, that is, any α such that $\alpha > \alpha_1 > \alpha_0$, we have

$$|f(t)| < M_1 e^{\alpha t}$$

Moreover, since $f(t)$ is piecewise regular, it is bounded over the finite interval $0 \leqq t \leqq T$; that is, there exists a positive number M_2 such that

$$|f(t)| < M_2 \equiv (M_2 e^{-\alpha t})e^{\alpha t} \qquad \text{for } 0 \leqq t \leqq T$$

Thus if we let M be the largest of the three numbers M_1, M_2, $M_2 e^{-\alpha T}$,† it is clear that

$$|f(t)| < Me^{\alpha t} \qquad \text{for } all\ t \geqq 0$$

Hence, returning to the integral in (4) and replacing $f(t)$ by its upper bound, we have

$$I \equiv \int_0^b |f(t)|e^{-st}\,dt \leqq \int_0^b Me^{\alpha t}e^{-st}\,dt = \left.\frac{Me^{-(s-\alpha)t}}{-(s-\alpha)}\right|_0^b = \frac{M}{s-\alpha}\left(1 - e^{-(s-\alpha)b}\right)$$

Now if $s > \alpha$, the last expression increases monotonically and approaches $M/(s - \alpha)$ as b becomes infinite. Therefore,

$$I \leqq \frac{M}{s-\alpha} \qquad s > \alpha > \alpha_0$$

Since the integrand of I is everywhere nonnegative, it is clear that I is a monotonically increasing function of b. Hence, being bounded above, as we have just shown, it must approach a limit as b becomes infinite. Since $s > \alpha > \alpha_0$ is clearly equivalent to the condition $s > \alpha_0$, the theorem is established.

Since the absolute value of an integral is always equal to or less than the integral of the absolute value, it follows from the preceding discussion that

$$\left|\int_0^b f(t)e^{-st}\,dt\right| \leqq \int_0^b |f(t)|e^{-st}\,dt \equiv I \leqq \frac{M}{s-\alpha}$$

Hence, letting $b \to \infty$, we have the important result

THEOREM 5

If $f(t)$ is piecewise regular and of exponential order with abscissa of convergence α_0, then, for all values of s and α such that $s > \alpha > \alpha_0$,

$$|\mathcal{L}\{f(t)\}| \leqq \frac{M}{s-\alpha} \qquad \text{where } M \text{ is independent of } s$$

Finally, from Theorem 5 we draw the following interesting conclusions

COROLLARY 1

If $f(t)$ is piecewise regular and of exponential order, then $\mathcal{L}\{f(t)\}$ approaches zero as s becomes infinite.

---

† Both M_1 and $M_2 e^{-\alpha T}$ must be considered, because if $\alpha > 0$ then M_2 is the maximum of $M_2 e^{-\alpha t}$ on $0 \leqq t \leqq T$, whereas if $\alpha < 0$ then $M_2 e^{-\alpha T}$ is the maximum of $M_2 e^{-\alpha t}$ on $0 \leqq t \leqq T$

COROLLARY 2

If $f(t)$ is piecewise regular and of exponential order, then $s\mathcal{L}\{f(t)\}$ is bounded as s becomes infinite

These corollaries make it clear that not all functions of s are Laplace transforms—or at least not Laplace transforms of functions of the "respectable" class defined by conditions **a** and **b′** For instance, $\phi(s) = s/(s - 1)$ does not approach zero as s becomes infinite, hence it is not the Laplace transform of any "respectable" function Also, although $\phi(s) = 1/\sqrt{s}$ does approach zero as s becomes infinite, it is not the transform of any "respectable" function, since $s\phi(s) \equiv \sqrt{s}$ is not bounded as s becomes infinite

We are now in a position to establish the uniform convergence of the integral defining $\mathcal{L}\{f(t)\}$

THEOREM 6

If $f(t)$ is piecewise regular and of exponential order with abscissa of convergence α_0, then, for any number $s_0 > \alpha_0$,

$$\mathcal{L}\{f(t)\} = \int_0^\infty f(t)e^{-st}\,dt$$

converges uniformly for all values of s such that $s \geqq s_0$

PROOF To prove this theorem, we must show that, given any $\epsilon > 0$, there exists a number B, depending on ϵ but not on s, such that

$$\left| \int_b^\infty f(t)e^{-st}\,dt \right| < \epsilon \qquad \text{for all } b > B \text{ and all } s \geqq s_0$$

Now
$$\left| \int_b^\infty f(t)e^{-st}\,dt \right| \leqq \int_b^\infty |f(t)|e^{-st}\,dt$$

and we know that for $s > \alpha_0$ the integral on the right approaches zero as b becomes infinite, since this is implied by the fact that

$$\int_0^\infty |f(t)|e^{-st}\,dt$$

is convergent for $s > \alpha_0$ (Theorem 1) In other words, given any $\epsilon > 0$ and any $s_0 > \alpha_0$, there exists a number B such that

$$\int_b^\infty |f(t)|e^{-s_0t}\,dt < \epsilon \qquad \text{for all } b > B$$

Now if $s \geqq s_0$, it is obvious that $e^{-st} \leqq e^{-s_0t}$ Hence,

$$\int_b^\infty |f(t)|e^{-st}\,dt \leqq \int_b^\infty |f(t)|e^{-s_0t}\,dt$$

and so for any $s \geqq s_0$ the integral on the left is less than ϵ for all values of b greater than the particular B which suffices for the integral on the right This value of B is clearly independent of s, and so the proof of the theorem is complete

In succeeding sections we shall find that many relatively complicated operations upon $f(t)$, such as differentiation and integration, for instance, can be replaced by simple algebraic operations such as multiplication or division by s, upon the transform of $f(t)$. This is analogous to the way in which such operations as multiplication and division of numbers are replaced by the simpler processes of addition and subtraction when we work not with the numbers themselves, but with their logarithms Our primary purpose in this chapter is to develop rules of transformation and tables of transforms which can be used, like tables of logarithms, to facilitate the manipulation of functions and by means of which we can recover the proper function from its transform at the end of a problem

EXERCISES

1 Prove that a function $f(t)$ is of exponential order if and only if s can be chosen so that $\lim_{t\to\infty} e^{-st}f(t) = 0$ If $f(t)$ is of exponential order, show that its abscissa of convergence α_0 is the greatest lower bound of all values of s such that $\lim_{t\to\infty} e^{-st}f(t) = 0$

2 Which of the following functions are of exponential order (a) t^n, (b) $\tan t$, (c) e^{t^2}, (d) $\cosh t$, (e) $1/t$, (f) t^2e^{3t}?

3 Prove that, if a piecewise regular function satisfies condition **b′**, it also satisfies condition **b** (Hint The proof of this is very much like the proof of Theorem 4)

4 Prove that, if a piecewise regular function satisfies condition **b**, it does not necessarily satisfy condition **b′** Hint Consider the function

$$f(t) = \begin{cases} e^{t^2} & t = n \\ 0 & t \neq n \end{cases} \qquad n = 0, 1, 2, 3,$$

5 Prove that, if $f(t)$ is piecewise regular and of exponential order, then $\int_0^t f(t)\,dt$ is also piecewise regular and of exponential order Show also that, if α_0 and α_1 are, respectively, the abscissas of convergence of $f(t)$ and $\int_0^t f(t)\,dt$ and if $\alpha_0 \geqq 0$, then $\alpha_1 \leqq \alpha_0$ Is it necessarily true that $\alpha_1 \leqq \alpha_0$ if $\alpha_0 < 0$?

7.2

The general method

The utility of the Laplace transformation is based primarily upon the following three theorems

THEOREM 1

$$\mathcal{L}\{c_1f_1(t) + c_2f_2(t)\} = c_1\mathcal{L}\{f_1(t)\} + c_2\mathcal{L}\{f_2(t)\}$$

PROOF To prove this, we have by definition

$$\begin{aligned}\mathcal{L}\{c_1f_1(t) + c_2f_2(t)\} &= \int_0^\infty [c_1f_1(t) + c_2f_2(t)]e^{-st}\,dt \\ &= c_1\int_0^\infty f_1(t)e^{-st}\,dt + c_2\int_0^\infty f_2(t)e^{-st}\,dt \\ &= c_1\mathcal{L}\{f_1(t)\} + c_2\mathcal{L}\{f_2(t)\} \qquad \text{as asserted}\end{aligned}$$

The extension of this theorem to linear combinations of more than two functions is obvious.

THEOREM 2

If $f(t)$ is a continuous, piecewise regular function of exponential order whose derivative is also piecewise regular and of exponential order and if $f(t)$ approaches the limit $f(0^+)$ as t approaches zero from the right, then the Laplace transform of $f'(t)$ is given by the formula

$$\mathcal{L}\{f'(t)\} = s\mathcal{L}\{f(t)\} - f(0^+)$$

provided s is greater than the abscissa of convergence of $f(t)$.

PROOF To prove this, let us suppose for definiteness that there is a single point, say $t = t_0$, where, though $f(t)$ is itself continuous, its derivative has a finite jump, as suggested by Fig. 7.1. Then, by definition

$$\begin{aligned}\mathcal{L}\{f'(t)\} &= \int_0^\infty f'(t)e^{-st}\,dt\\ &= \lim_{\substack{\delta_1,\delta_2,\delta_3\to 0\\ b\to\infty}} \left[\int_{\delta_1}^{t_0-\delta_2} f'(t)e^{-st}\,dt + \int_{t_0+\delta_3}^{b} f'(t)e^{-st}\,dt\right]\end{aligned}$$

If we use integration by parts on these integrals, choosing

$$\begin{aligned} u &= e^{-st} & dv &= f'(t)\,dt\\ du &= -se^{-st}\,dt & v &= f(t)\end{aligned}$$

we have

$$\begin{aligned}\mathcal{L}\{f'(t)\} = \lim_{\substack{\delta_1,\delta_2,\delta_3\to 0\\ b\to\infty}} \Bigg[e^{-st}f(t)\Big|_{\delta_1}^{t_0-\delta_2} &+ s\int_{\delta_1}^{t_0-\delta_2} f(t)e^{-st}\,dt\\ &+ e^{-st}f(t)\Big|_{t_0+\delta_3}^{b} + s\int_{t_0+\delta_3}^{b} f(t)e^{-st}\,dt\Bigg]\end{aligned}$$

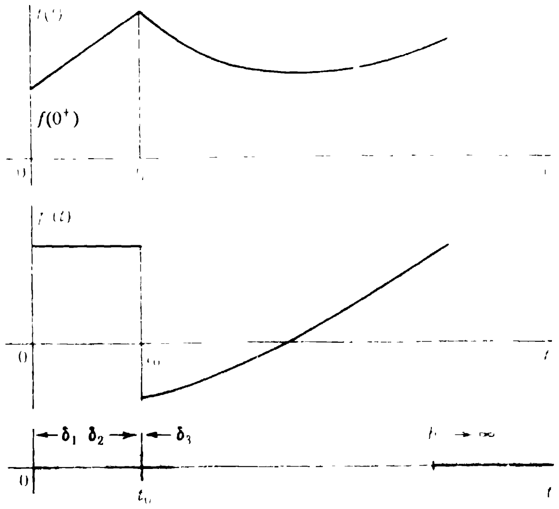


FIGURE 7.1
A continuous function whose derivative has a point of discontinuity

In the limit the two integrals which remain combine to give precisely

$$s\int_0^\infty f(t)e^{-st}\,dt = s\mathcal{L}\{f(t)\}$$

Similarly, the first evaluated portion yields

$$e^{-st_0}f(t_{0-}) - f(0^+)$$

and the second yields simply

$$-e^{-st_0}f(t_{0+})$$

because, since $f(t)$ is of exponential order, s can be chosen sufficiently large [i.e., greater than the abscissa of convergence of $f(t)$] that the contribution from the upper limit is zero. Now $f(t)$ was assumed to be continuous. Hence at t_0 (as at all other points) its right- and left-hand limits must be equal. Therefore, the terms

$$e^{-st_0}f(t_{0-}) \qquad \text{and} \qquad -e^{-st_0}f(t_{0+})$$

cancel, leaving finally

$$\mathcal{L}\{f'(t)\} = s\mathcal{L}\{f(t)\} - f(0^+) \qquad \text{as asserted}$$

The extension of the preceding proof to functions whose derivatives have more than one finite jump is obvious. The extension of the theorem to the relatively unimportant case in which $f(t)$ itself is permitted to have finite jumps is indicated in Exercise 3.

COROLLARY 1

If both $f(t)$ and $f'(t)$ are continuous, piecewise regular functions of exponential order and if $f''(t)$ is piecewise regular and of exponential order, then

$$\mathcal{L}\{f''(t)\} = s^2\mathcal{L}\{f(t)\} - sf(0^+) - f'(0^+)$$

where $f(0^+)$ and $f'(0^+)$ are, respectively, the values that $f(t)$ and $f'(t)$ approach as t approaches zero from the right.

PROOF This result follows immediately by applying Theorem 2 twice to $f''(t)$:

$$\begin{aligned}\mathcal{L}\{f''(t)\} = \mathcal{L}\{[f'(t)]'\} &= s\mathcal{L}\{f'(t)\} - f'(0^+)\\ &= s[s\mathcal{L}\{f(t)\} - f(0^+)] - f'(0^+)\\ &= s^2\mathcal{L}\{f(t)\} - sf(0^+) - f'(0^+) \text{ as asserted}\end{aligned}$$

The extension of this result to derivatives of higher order is obvious (Exercise 1).

THEOREM 3

If $f(t)$ is piecewise regular and of exponential order, then the Laplace transform of $\int_a^t f(t)\,dt$ is given by the formula

$$\mathcal{L}\left\{\int_a^t f(t)\,dt\right\} = \frac{1}{s}\mathcal{L}\{f(t)\} + \frac{1}{s}\int_a^0 f(t)\,dt$$

PROOF To prove this theorem, we have by definition

$$\mathcal{L}\left\{\int_a^t f(t)\,dt\right\} = \int_0^\infty \left[\int_a^t f(x)\,dx\right] e^{-st}\,dt$$

where the dummy variable x has been introduced for convenience. If we integrate the last integral by parts, with

$$u = \int_a^t f(x)\,dx \qquad dv = e^{-st}\,dt$$

$$du = f(t)\,dt \qquad v = \frac{e^{-st}}{-s}$$

we have $$\mathcal{L}\left\{\int_a^t f(t)\,dt\right\} = \left[\frac{e^{-st}}{-s}\int_a^t f(x)\,dx\right]_0^\infty + \frac{1}{s}\int_0^\infty f(t)e^{-st}\,dt$$

Since $f(t)$ is of exponential order, so, too, is its integral (Exercise 5, Sec. 7.1). Hence s can be chosen sufficiently large that the integrated portion vanishes at the upper limit, leaving

$$\mathcal{L}\left\{\int_a^t f(t)\,dt\right\} = \frac{1}{s}\int_a^0 f(x)\,dx + \frac{1}{s}\mathcal{L}\{f(t)\} \qquad \text{as asserted}$$

The extension of this result to repeated integrals of $f(t)$ is obvious (Exercise 2).

Although we need many more formulas before the Laplace transformation can be applied effectively to specific problems, Theorems 1, 2, and 3 allow us to outline all the essential steps in the usual application of this method to the solution of differential equations. Suppose that we are given the equation

$$ay'' + by' + cy = f(t)$$

If we take the Laplace transform of both sides, we have by Theorem 1

$$a\mathcal{L}\{y''\} + b\mathcal{L}\{y'\} + c\mathcal{L}\{y\} = \mathcal{L}\{f(t)\}$$

Now applying Theorem 2 and its corollary, we have

$$a[s^2\mathcal{L}\{y\} - sy_0 - y'_0] + b[s\mathcal{L}\{y\} - y_0] + c\mathcal{L}\{y\} = \mathcal{L}\{f(t)\}$$

where y_0 and y'_0 are the given initial values of y and y'. Collecting terms on $\mathcal{L}\{y\}$ and then solving for $\mathcal{L}\{y\}$, we obtain finally

$$\mathcal{L}\{y\} = \frac{\mathcal{L}\{f(t)\} + (as + b)y_0 + ay'_0}{as^2 + bs + c}$$

Now $f(t)$ is a given function of t, hence its Laplace transform (if it exists) is a perfectly definite function of s (although as yet we have no specific formulas for finding it). Moreover, y_0 and y'_0 are definite numbers, known from the data of the problem. Hence the transform of y is a completely known function of s. Thus if we had available a table of transforms, we could find in it the function $y(t)$ having the right-hand side of the last equation for its transform, *and this function would be the formal solution to our problem, initial conditions and all.* The formal solution could then be substituted into the given differential equation to verify that it was indeed the genuine solution.

This brief discussion illustrates the two great advantages of the Laplace transformation in solving linear, constant-coefficient differential equations: first, the way in which it reduces the problem to one in algebra; second, the automatic way in which it takes care of initial conditions without the necessity of constructing a general solution and then specializing the arbitrary constants it contains Clearly, our immediate task is to implement this process by establishing an adequate table of transforms

EXERCISES

1 Show that $\mathcal{L}\{f'''\} = s^3\mathcal{L}\{f\} - s^2f_0 - sf_0' - f_0''$ What is $\mathcal{L}\{f^{(n)}\}$?

2 Show that

$$\mathcal{L}\left\{\int_a^t\int_a^t f(t)\,dt\,dt\right\} = \frac{1}{s^2}\mathcal{L}\{f\} + \frac{1}{s^2}\int_a^0 f(t)\,dt + \frac{1}{s}\int_a^0\int_a^t f(t)\,dt\,dt$$

3 If $f(t)$ satisfies all the conditions of Theorem 2 except that it has an upward jump of magnitude J_0 at $t = t_0$, show that

$$\mathcal{L}\{f'(t)\} = s\mathcal{L}\{f\} - f_0 - J_0e^{-st_0}$$

4 Show that

$$\mathcal{L}\{f(at)\} = \frac{1}{a}\mathcal{L}\{f(t)\}\Big|_{s\to\frac{s}{a}}$$

5 a Given $\mathcal{L}\{\cos t\} = s/(s^2+1)$, use the result of Exercise 4 to determine $\mathcal{L}\{\cos at\}$

b Given $\mathcal{L}\{\sin t\} = 1/(s^2+1)$, use the result of Exercise 4 to determine $\mathcal{L}\{\sin at\}$

6 Explain how the Laplace transform can be used to solve a system of simultaneous linear differential equations with constant coefficients In particular, given that $y = y_0$ and $z = z_0$ when $t = 0$, obtain formulas for the Laplace transforms of y and z if

$$a_1\frac{dy}{dt} + b_1y + c_1\frac{dz}{dt} + d_1z = f_1(t)$$

$$a_2\frac{dy}{dt} + b_2y + c_2\frac{dz}{dt} + d_2z = f_2(t)$$

7 The function

$$S[f(t)] \equiv \int_0^\pi f(t)\sin nt\,dt \qquad n = 1, 2, 3,$$

is called the **sine transform** of $f(t)$ Show that

$$S(f'') = -n^2S(f) + n[f(0) - (-1)^nf(\pi)]$$

8 The function

$$C[f(t)] \equiv \int_0^\pi f(t)\cos nt\,dt \qquad n = 0, 1, 2,$$

is called the **cosine transform** of $f(t)$ Obtain a formula expressing $C(f'')$ in terms of $C(f)$

9 Let $T[f(t)]$ be a general integral transform

$$T[f(t)] \equiv \int_a^b f(t)K(s,t)\,dt$$

where $K(s,t)$ is the so-called **kernel** of the transformation Obtain conditions on $K(s,t)$ so that $T(f')$ and $T(f'')$ contain no terms involving the evaluation of f or any of its derivatives Find at least one kernel satisfying these conditions.

10 If $f''(t)$ and $f'(t)$ are both piecewise regular and of exponential order and if $f(t)$ is continuous and $f(0^+) = 0$, show that as s becomes infinite $\mathcal{L}\{f(t)\}$ tends to zero at least as rapidly as $1/s^2$ Can this result be generalized?

7.3

The transforms of special functions

Among all the functions whose transforms we might now think of tabulating, the most important are the simple ones

$$e^{-at} \qquad \cos bt \qquad \sin bt \qquad t^n$$

and the unit step function,

$$u(t) = \begin{cases} 0 & t < 0 \\ 1 & t > 0 \end{cases}$$

shown in Fig. 7.2. Once we know the transforms of these functions, nearly all the formulas we shall need can be obtained through the use of a few additional general theorems which we shall establish in the next section. The specific results are the following:

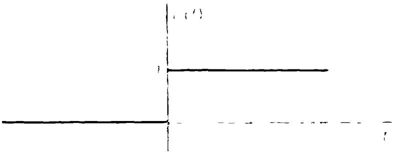


FIGURE 7.2
The unit step function $u(t)$

FORMULA 1

$$\mathcal{L}\{e^{-at}\} = \frac{1}{s+a}$$

FORMULA 2

$$\mathcal{L}\{\cos bt\} = \frac{s}{s^2+b^2}$$

FORMULA 3

$$\mathcal{L}\{\sin bt\} = \frac{b}{s^2+b^2}$$

FORMULA 4

$$\mathcal{L}\{t^n\} = \begin{cases} \dfrac{\Gamma(n+1)}{s^{n+1}} & n > -1 \\[2ex] \dfrac{n!}{s^{n+1}} & n \text{ a positive integer} \end{cases}$$

FORMULA 5

$$\mathcal{L}\{u(t)\} = \frac{1}{s}$$

To prove Formula 1 we have simply

$$\mathcal{L}\{e^{-at}\} = \int_0^\infty e^{-at}e^{-st}\,dt = \frac{e^{-(s+a)t}}{-(s+a)}\bigg|_0^\infty = \frac{1}{s+a} \qquad \text{if } s+a > 0$$

To prove Formula 2, we have

$$\mathcal{L}\{\cos bt\} = \int_0^\infty \cos bt\, e^{-st}\,dt = \frac{e^{-st}}{s^2+b^2}(-s\cos bt + b\sin bt)\Big|_0^\infty$$

$$= \frac{s}{s^2+b^2} \qquad \text{if } s > 0$$

To prove Formula 3, we have

$$\mathcal{L}\{\sin bt\} = \int_0^\infty \sin bt\, e^{-st}\,dt = \frac{e^{-st}}{s^2+b^2}(-s\sin bt - b\cos bt)\Big|_0^\infty$$

$$= \frac{b}{s^2+b^2} \qquad \text{if } s > 0$$

Before we can prove Formula 4 it will be necessary for us to investigate briefly the so-called **gamma** or **generalized factorial function** defined by the equation

(1) $$\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}\,dt$$

This improper integral can be shown to be convergent for all $x > 0$

To determine the simple properties of the gamma function and its relation to the familiar factorial function

$$n! = n(n-1) \cdots 3 \cdot 2 \cdot 1$$

defined in elementary algebra for positive integral values of n, let us apply integration by parts to the definitive integral (1), taking

$$u = e^{-t} \qquad dv = t^{x-1}\,dt \qquad du = -e^{-t}\,dt \qquad v = \frac{t^x}{x}$$

Then $$\Gamma(x) = \frac{t^x e^{-t}}{x}\Big|_0^\infty + \frac{1}{x}\int_0^\infty e^{-t}t^x\,dt$$

Under the restriction $x > 0$, the integrated portion vanishes at both limits By comparison with (1), it is clear that the integral which remains is simply $\Gamma(x+1)$ Thus we have established the important recurrence relation

(2) $$\Gamma(x) = \frac{\Gamma(x+1)}{x} \qquad x > 0$$

or

(2a) $$x\Gamma(x) = \Gamma(x+1)$$

Moreover, we have specifically

$$\Gamma(1) = \int_0^\infty e^{-t}\,dt = -e^{-t}\Big|_0^\infty = 1$$

Therefore, using (2a),

$$\Gamma(2) = 1 \cdot \Gamma(1) = 1$$

$$\Gamma(3) = 2 \cdot \Gamma(2) = 2 \cdot 1 = 2!$$

$$\Gamma(4) = 3 \cdot \Gamma(3) = 3 \cdot 2! = 3!$$

and in general

(3) $$\Gamma(n+1) = n! \qquad n = 1, 2, 3, \ldots$$

The connection between the gamma function and ordinary factorials is now clear. However, the gamma function constitutes an essential extension of the idea of a factorial, since its argument x is not restricted to positive integral values but can vary continuously over any interval which does not contain a nonnegative integer.

From (2) and the fact that $\Gamma(1) = 1$, it is evident that $\Gamma(x)$ becomes infinite as x approaches zero. It is thus clear that $\Gamma(x)$ cannot be defined for $x = 0, -1, -2, \ldots$ in a way consistent with Eq. (2), hence we shall leave it undefined for these values of x. For all other values of x, however, $\Gamma(x)$ is well defined, the use of the recurrence formula (2a) effectively removing the restriction that x be positive, which the integral definition (1) requires. By methods which need not concern us here, tables of $\Gamma(x)$ have been constructed and can be found, usually as tables of $\log \Gamma(x)$, in most elementary handbooks. Because of the recurrence formula which the gamma function satisfies, these tables ordinarily cover only a unit interval on x, usually the interval $1 \leqq x \leqq 2$. A plot of $\Gamma(x)$ is shown in Fig. 7.3.

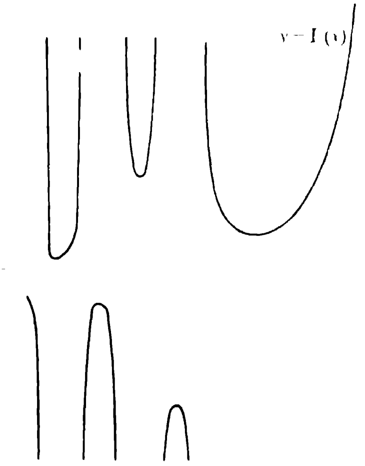


FIGURE 7.3 Plot of the function $y = \Gamma(x)$

EXAMPLE 1

What is the value of $I = \int_0^\infty \sqrt{z}\, e^{-z^3}\, dz$?

This integral is typical of many which can be reduced to the standard form of the gamma function by a suitable substitution. In this case it is clear on comparing the given integral with (1) that we should let

$$z = t^{1/3} \qquad dz = \tfrac{1}{3} t^{-2/3}\, dt$$

getting

$$I = \int_0^\infty \sqrt{t^{1/3}}\, e^{-t} \left(\tfrac{1}{3} t^{-2/3}\, dt\right) = \tfrac{1}{3} \int_0^\infty t^{1/2-1} e^{-t}\, dt = \tfrac{1}{3} \Gamma(\tfrac{1}{2})$$

Since $\Gamma(\tfrac{1}{2})$ cannot be found in the usual table, which lists $\Gamma(x)$ only for $1 \leqq x \leqq 2$, it is necessary to use the recurrence relation (2) to bring the argument of the gamma function into this interval:

$$\frac{1}{3}\Gamma\left(\frac{1}{2}\right) = \frac{1}{3}\,\frac{\Gamma(3/2)}{1/2} = \frac{2}{3}(0.88623) = 0.59082\dagger$$

Returning now to Formula 4, we have

$$\mathcal{L}\{t^n\} = \int_0^\infty t^n e^{-st}\,dt$$

In an attempt to reduce this to the standard form of the gamma function, let us make the substitution

$$t = \frac{z}{s} \qquad dt = \frac{dz}{s}$$

Then $$\mathcal{L}\{t^n\} = \int_0^\infty \left(\frac{z}{s}\right)^n e^{-z}\frac{dz}{s} = \frac{1}{s^{n+1}}\int_0^\infty z^n e^{-z}\,dz = \frac{\Gamma(n+1)}{s^{n+1}}$$

Since $\Gamma(n+1) = n!$ when n is a positive integer, this establishes the second part of Formula 4 also

It is interesting to note that, if n is negative,

$$s\mathcal{L}\{t^n\} = \frac{\Gamma(n+1)}{s^n}$$

is not bounded as $s \to \infty$ Hence, according to Corollary 2, Theorem 5, Sec. 7 1, this function of s is not the Laplace transform of a piecewise regular function of exponential order This, of course, is obvious, since when n is negative, t^n, though of exponential order (with abscissa of convergence $\alpha_0 = 0$), is not bounded in the neighborhood of the origin and so is not piecewise regular It can be shown, however, that the improper integral defining $\mathcal{L}\{t^n\}$ exists for $n > -1$ although it does not exist for $n \leqq -1$ Formula 4 must therefore be qualified by the restriction $n > -1$.

Formula 5 can be obtained immediately by taking $n = 0$ in Formula 4.

EXAMPLE 2

What is the Laplace transform of sinh bt?

Since sinh $bt = (e^{bt} - e^{-bt})/2$, we have

$$\mathcal{L}\{\sinh bt\} = \mathcal{L}\left\{\frac{e^{bt} - e^{-bt}}{2}\right\} = \frac{1}{2}\left(\frac{1}{s-b} - \frac{1}{s+b}\right) = \frac{b}{s^2 - b^2}$$

The analogy with Formula 3 for the transform of sin bt is apparent

EXAMPLE 3

If $\mathcal{L}\{y(t)\} = (s+1)/(s^2+s-6)$, what is $y(t)$?

None of our formulas yields a transform resembling this one However, using the method of partial fractions, we can write

$$\frac{s+1}{s^2+s-6} = \frac{s+1}{(s-2)(s+3)} = \frac{A}{s-2} + \frac{B}{s+3} = \frac{A(s+3)+B(s-2)}{(s-2)(s+3)}$$

---

† Actually the value of $\Gamma(1/2)$ is known exactly and in fact is equal to $\sqrt{\pi}$ (Exercise 10) Hence, in this example $I = \sqrt{\pi}/3$

For this to be an identity we must have

$$s + 1 = A(s + 3) + B(s - 2)$$

Setting $s = 2$ and $s = -3$ in turn, we find from this that $A = \frac{3}{5}$, $B = \frac{2}{5}$ Hence

$$\mathcal{L}\{y(t)\} = \frac{1}{5}\left(\frac{3}{s - 2} + \frac{2}{s + 3}\right)$$

Formula 1 can now be applied to the individual terms, and we find

$$y(t) = \tfrac{1}{5}(3e^{2t} + 2e^{-3t})$$

EXAMPLE 4

Solve for $y(t)$ from the simultaneous equations

$$y' + 2y + 6\int_0^t z\,dt = -2u(t)$$

$$y' + z' + z = 0$$

if $y_0 = -5$ and $z_0 = 6$

We begin by taking the Laplace transform of each equation term by term

$$[s\mathcal{L}\{y\} + 5] + 2\mathcal{L}\{y\} + \frac{6}{s}\mathcal{L}\{z\} = -\frac{2}{s}$$

$$[s\mathcal{L}\{y\} + 5] + [s\mathcal{L}\{z\} - 6] + \mathcal{L}\{z\} = 0$$

Obvious simplifications then lead to the following pair of linear algebraic equations in the transforms of the unknown functions $y(t)$ and $z(t)$

$$(s^2 + 2s)\mathcal{L}\{y\} + 6\mathcal{L}\{z\} = -2 - 5s$$

$$s\mathcal{L}\{y\} + (s + 1)\mathcal{L}\{z\} = 1$$

Since it is $y(t)$ that we are asked to find, we solve these simultaneous equations for $\mathcal{L}\{y\}$, getting

$$\mathcal{L}\{y\} = \frac{\begin{vmatrix} -2 - 5s & 6 \\ 1 & s + 1 \end{vmatrix}}{\begin{vmatrix} s^2 + 2s & 6 \\ s & s + 1 \end{vmatrix}} = \frac{-5s^2 - 7s - 8}{s^3 + 3s^2 - 4s}$$

Applying the method of partial fractions to this expression, we have

$$\mathcal{L}\{y\} = \frac{-5s^2 - 7s - 8}{s^3 + 3s^2 - 4s} = \frac{2}{s} - \frac{4}{s - 1} - \frac{3}{s + 4}$$

Finally, taking the inverse of each of these terms, we find

$$y(t) = 2u(t) - 4e^{t} - 3e^{-4t}$$

EXERCISES

1 What is $\mathcal{L}\{\cosh bt\}$?

2 What is $\mathcal{L}\{\cos(at + b)\}$? [Hint First express $\cos(at + b)$ as the difference of two terms]

3 What is $\mathcal{L}\{\cos^2 bt\}$? (Hint First express $\cos^2 bt$ as a function of $2bt$.)

4 What is $\mathcal{L}\{(t + 1)^2\}$?

5 Find the inverse of each of the following functions:

a $\dfrac{1}{s + 3}$ b $\dfrac{1}{s^4}$ c $\dfrac{1}{s^2 + 9}$ d $\dfrac{2s + 3}{s^2 + 9}$ e $\dfrac{s + 3}{(s + 1)(s - 3)}$

6 Find the solution of each of the following differential equations:

a $y'' + 4y' - 5y = 0$ $y_0 = 1,\ y'_0 = 0$

b $y'' - 4y = 0$ $y_0 = -1,\ y'_0 = 1$

c $4y'' + y = 0$ $\qquad y_0 = 2,\ y_0' = 1$

d $y'' + 4y = u(t)$ $\qquad y_0 = y_0' = 0$

e $y'' + 3y' + 2y = e^t$ $\qquad y_0 = 1,\ y_0' = 0$

f $y''' + 6y'' + 11y' + 6y = 0$ $\qquad y_0 = 2,\ y_0' = 1,\ y_0'' = -1$

7 Find the solution of the following system of equations:

$$\begin{aligned} y' + y + 2z' + 3z &= e^{-t} \\ 3y' - y + 4z' + z &= 0 \end{aligned} \qquad y_0 = 0,\ z_0 = 1$$

8 Find the solution of the following system of equations:

$$\begin{aligned} (D + 1)y + (2D + 3)z &= 0 \\ (D - 4)y + (3D - 8)z &= \sin t \end{aligned} \qquad y_0 = -1,\ z_0 = 0$$

9 Evaluate each of the following integrals

a $\displaystyle\int_0^\infty \frac{e^{-x}}{\sqrt{x}}\,dx$ $\qquad$ b $\displaystyle\int_0^\infty e^{-\sqrt[3]{x}}\,dx$ $\qquad$ c $\displaystyle\int_0^\infty (x + 1)^2 e^{-x^2}\,dx$

d $\displaystyle\int_0^\infty \frac{x^c}{c^x}\,dx$ (Hint. $c^x = e^{x \ln c}$)

10 Show that $\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$ [Hint First show $\Gamma(\tfrac{1}{2}) = 2\int_0^\infty e^{-x^2}\,dx = 2\int_0^\infty e^{-y^2}\,dy$ Then multiply these integrals and evaluate the resulting repeated integral by changing to polar coordinates]

7.4

Further general theorems

We are now in a position to derive a number of theorems that will be of considerable use in the application of the Laplace transformation to practical problems We begin with a result which allows us to infer the behavior of a function $f(t)$ for small positive values of t from the behavior of $\mathcal{L}\{f(t)\}$ for large positive values of s

THEOREM 1

If $f(t)$ and $f'(t)$ are both piecewise regular and of exponential order, then

$$\lim_{s\to\infty} s\mathcal{L}\{f(t)\} = \lim_{t\to 0^+} f(t) = f(0^+)$$

PROOF For convenience we shall prove this under the additional assumption that $f(t)$ is continuous, leaving as an exercise the proof under the less restrictive conditions of the theorem as stated. We may thus begin with the result of Theorem 2, Sec 7.2, namely,

$$\mathcal{L}\{f'(t)\} = s\mathcal{L}\{f(t)\} - f(0^+)$$

Hence, taking the limit of each side,

$$(1) \qquad \lim_{s\to\infty} \mathcal{L}\{f'(t)\} = \lim_{s\to\infty} s\mathcal{L}\{f(t)\} - f(0^+)$$

However, under the conditions of the theorem, it follows from Corollary 1, Theorem 5, Sec. 7.1, that

$$\lim_{s\to\infty} \mathcal{L}\{f'(t)\} = 0$$

Therefore, from (1),

$$\lim_{s\to\infty} s\mathcal{L}\{f(t)\} = f(0^+) \qquad \text{as asserted.}$$

An analogous result which allows us to infer the behavior of a function $f(t)$ for large positive values of t from the behavior of $\mathcal{L}\{f(t)\}$ for small values of s is contained in the following theorem.

THEOREM 2

If $f(t)$ and $f'(t)$ are both piecewise regular and of exponential order and if the abscissa of convergence of $f'(t)$ is negative, then

$$\lim_{s\to 0} s\mathcal{L}\{f(t)\} = \lim_{t\to +\infty} f(t)$$

provided these limits exist.

PROOF Here, as in the proof of Theorem 1, we shall base our argument on the additional assumption that $f(t)$ is continuous. Then again we may take limits in the result of Theorem 2, Sec. 7.2, getting

$$\lim_{s\to 0} \mathcal{L}\{f'(t)\} = \lim_{s\to 0} s\mathcal{L}\{f(t)\} - f(0^+)$$

or

$$(2) \qquad \lim_{s\to 0} s\mathcal{L}\{f(t)\} = \lim_{s\to 0} \mathcal{L}\{f'(t)\} + f(0^+)$$

But

$$\lim_{s\to 0} \mathcal{L}\{f'(t)\} = \lim_{s\to 0} \int_0^\infty f'(t)e^{-st}\,dt$$

and under the conditions of the present theorem we can invoke Theorems 6 and 1, Sec. 7.1, and take the limit on the right inside the integral sign. Thus

$$\begin{aligned}\lim_{s\to 0} \mathcal{L}\{f'(t)\} &= \int_0^\infty f'(t)\left(\lim_{s\to 0} e^{-st}\right) dt = \int_0^\infty f'(t)\,dt = f(t)\Big|_0^\infty \\ &= \lim_{t\to\infty} f(t) - f(0^+)\end{aligned}$$

Substituting this into (2) we have finally

$$\begin{aligned}\lim_{s\to 0} s\mathcal{L}\{f(t)\} &= \left[\lim_{t\to\infty} f(t) - f(0^+)\right] + f(0^+) \\ &= \lim_{t\to\infty} f(t) \qquad \text{as asserted.*}\end{aligned}$$

---

* In realistic applications of this theorem, $\mathcal{L}\{f(t)\}$ will be known, but $f(t)$ and its abscissa of convergence will be unknown. Hence it is desirable that conditions for the use of the theorem be expressed in terms of $\mathcal{L}\{f(t)\}$ rather than $f(t)$. This can be done, since it is possible to show that Theorem 2 cannot be applied if there is any value of s with nonnegative real part for which $s\mathcal{L}\{f(t)\}$ is unbounded, but can be applied if no such value exists. For example, even though $\lim_{s\to 0} s/(s^2+1)$ exists, Theorem 2 cannot be applied to $\mathcal{L}\{f(t)\} = 1/(s^2+1)$, since this is unbounded for the values $s = \pm i \equiv 0 \pm i$. In this case, of course, $f(t) = \sin t$, and clearly $\lim_{t\to\infty} \sin t$ does not exist.

When the Laplace transform of an unknown function $f(t)$ contains the factor s,[†] it is often convenient to find $f(t)$ by means of the following theorem

THEOREM 3

If $f(t)$ is piecewise regular and of exponential order, if $\mathcal{L}\{f(t)\} = s\phi(s)$, and if the inverse of the factor $\phi(s)$ is continuous for $t > 0$, then

$$f(t) = \frac{d}{dt}\mathcal{L}^{-1}\{\phi(s)\}$$

PROOF To prove this theorem, let $F(t) \equiv \mathcal{L}^{-1}\{\phi(s)\}$ be the function which has $\phi(s)$ for its transform If $F(t)$ is continuous, as assumed, then, by Theorem 2, Sec 7 2,

$$\mathcal{L}\{F'(t)\} = s\mathcal{L}\{F(t)\} - F(0^+) = s\phi(s) - F(0^+)$$

But, by Theorem 1,

$$F(0^+) = \lim_{s\to\infty} s\mathcal{L}\{F(t)\} = \lim_{s\to\infty} s\phi(s) = 0$$

where the last step follows from Corollary 1, Theorem 5, Sec 7 1, since $s\phi(s)$ is the transform of the function $f(t)$, which, though unknown, is assumed to be "respectable." Hence,

$$\mathcal{L}\{f(t)\} = \mathcal{L}\{F'(t)\}$$

since each is equal to $s\phi(s)$ Therefore[‡]

$$f(t) = \frac{dF(t)}{dt} = \frac{d}{dt}\mathcal{L}^{-1}\{\phi(s)\} \qquad \text{as asserted}$$

EXAMPLE 1

What is $\mathcal{L}^{-1}\{s/(s^2 + 4)\}$?

By Formula 2, Sec 7 3, we see immediately that the required inverse is $f(t) = \cos 2t$ However, it is interesting that we can also obtain this result by suppressing the factor s, finding the inverse $F(t)$ of the remaining portion of the transform, namely,

$$\frac{1}{s^2 + 4}$$

and then differentiating this inverse according to Theorem 3

$$f(t) = \frac{d}{dt}\mathcal{L}^{-1}\left\{\frac{1}{s^2 + 4}\right\} = \frac{d}{dt}\left(\frac{\sin 2t}{2}\right) = \cos 2t$$

as before The usual applications of this theorem are, of course, not of this trivial character

---

[†] This can always be arranged, of course, by multiplying and dividing the transform by s, that is, $\phi(s) \equiv s[\phi(s)/s]$

[‡] This, of course, assumes the "obvious" theorem that, if two functions have the same transform, they are identical This is strictly true if the functions are continuous If discontinuities are permitted, the most we can say is that two functions with the same transform cannot differ over any interval of positive length, although they may differ at various isolated points A detailed discussion of this result (Lerch's theorem) would take us too far afield

When the Laplace transform of an unknown function $f(t)$ contains the factor $1/s$,† it is often convenient to find $f(t)$ by means of the following theorem:

THEOREM 4

If $f(t)$ *is piecewise regular and of exponential order and if* $\mathcal{L}\{f(t)\} = \phi(s)/s$, *then*

$$f(t) = \int_0^t \mathcal{L}^{-1}\{\phi(s)\}\, dt$$

PROOF To prove this theorem, let $F(t) \equiv \mathcal{L}^{-1}\{\phi(s)\}$ be the function which has $\phi(s)$ for its transform Then, by Theorem 3, Sec. 7 2,

$$\int_0^t F(t)\, dt = \frac{1}{s}\,\mathcal{L}\{F(t)\} + \frac{1}{s}\int_0^0 F(t)\, dt = \frac{1}{s}\,\mathcal{L}\{F(t)\} = \frac{\phi(s)}{s}$$

Thus both $f(t)$ and $\int_0^t F(t)\, dt \equiv \int_0^t \mathcal{L}^{-1}\{\phi(s)\}\, dt$ have $\phi(s)/s$ for their Laplace transform, and so must be equal, as asserted

EXAMPLE 2

What is $\mathcal{L}^{-1}\{1/s(s^2 + 4)\}$?

Here, using the last theorem, we first suppress the factor $1/s$, getting

$$\phi(s) = \frac{1}{s^2 + 4}$$

By Formula 3, Sec 7 3, the inverse of this is $F(t) = \frac{1}{2}\sin 2t$ Finally, we obtain $f(t)$ by integrating $F(t)$ from 0 to t

$$f(t) = \int_0^t \frac{\sin 2t}{2}\, dt = -\frac{\cos 2t}{4}\Bigg|_0^t = \frac{1 - \cos 2t}{4}$$

One of the most useful properties of the Laplace transformation is contained in the so-called **first shifting theorem**:

THEOREM 5

$$\mathcal{L}\{e^{-at}f(t)\} = \mathcal{L}\{f(t)\}_{s \to s+a}$$

PROOF By definition,

$$\mathcal{L}\{e^{-at}f(t)\} = \int_0^\infty [e^{-at}f(t)]e^{-st}\, dt = \int_0^\infty f(t)e^{-(s+a)t}\, dt$$

and the last integral is in structure exactly the Laplace transform of $f(t)$ itself, except that $s + a$ takes the place of s

In words, Theorem 5 says that *the transform of* e^{-at} *times a function of t is equal to the transform of the function itself, with s replaced by* $s + a$

As a tool for finding inverses, this theorem asserts that, if we reverse the substitution $s \to s + a$, that is, if we replace s by $s - a$, then the inverse of the modified transform $\phi(s - a)$ must be multiplied by e^{-at} to obtain the inverse of the original transform This procedure is summarized in the following result

---

† This can always be arranged, of course, by multiplying and dividing the transform by s, that is, $\phi(s) = (1/s)[s\phi(s)]$

COROLLARY 1

$$\mathcal{L}^{-1}\{\phi(s)\} = e^{-at}\mathcal{L}^{-1}\{\phi(s-a)\}$$

By means of Theorem 5 we can easily establish the following important formulas

FORMULA 1

$$\mathcal{L}\{e^{-at}\cos bt\} = \frac{s+a}{(s+a)^2+b^2}$$

FORMULA 2

$$\mathcal{L}\{e^{-at}\sin bt\} = \frac{b}{(s+a)^2+b^2}$$

FORMULA 3

$$\mathcal{L}\{e^{-at}t^n\} = \begin{cases} \dfrac{\Gamma(n+1)}{(s+a)^{n+1}} & n > -1 \\[2ex] \dfrac{n!}{(s+a)^{n+1}} & n \text{ a positive integer} \end{cases}$$

EXAMPLE 3

If $\mathcal{L}\{y\} = (2s+5)/(s^2+4s+13)$, what is y?

By obvious manipulations we obtain

$$\mathcal{L}\{y\} = \frac{2(s+2)+1}{(s+2)^2+3^2} = 2\left[\frac{s+2}{(s+2)^2+3^2}\right] + \frac{1}{3}\left[\frac{3}{(s+2)^2+3^2}\right]$$

Hence, by Formulas 1 and 2,

$$y = 2e^{-2t}\cos 3t + \tfrac{1}{3}e^{-2t}\sin 3t$$

EXAMPLE 4

What is the solution of the differential equation

$$y'' + 2y' + y = te^{-t}$$

for which $y_0 = 1$ and $y_0' = -2$?

Transforming both sides of the given equation, we have

$$(s^2\mathcal{L}\{y\} - s + 2) + 2(s\mathcal{L}\{y\} - 1) + \mathcal{L}\{y\} = \frac{1}{(s+1)^2}$$

$$(s^2 + 2s + 1)\mathcal{L}\{y\} = \frac{1}{(s+1)^2} + s$$

$$\mathcal{L}\{y\} = \frac{1}{(s+1)^4} + \frac{s}{(s+1)^2}$$

By Formula 3, the inverse of the first fraction in $\mathcal{L}\{y\}$ is $\frac{1}{3!}t^3e^{-t}$ To find the inverse of the second fraction we can write it in the form

$$\frac{s+1-1}{(s+1)^2} = \frac{1}{s+1} - \frac{1}{(s+1)^2}$$

and take the inverse of each term, or we can suppress the factor s, take the inverse of what remains, and differentiate this result. By either method we obtain immediately $e^{-t} - te^{-t}$. Hence

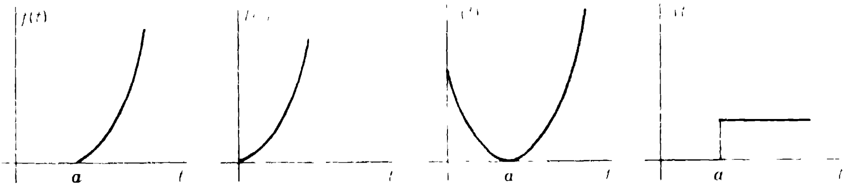
$$y = \frac{t^3 e^{-t}}{3!} + e^{-t} - te^{-t}$$

In this example the characteristic equation of the differential equation has repeated roots, and moreover the term on the right is a part of the complementary function, yet neither of these features requires any special treatment in the operational solution of the problem. This is another of the many advantages of the Laplace transform method of solving linear differential equations with constant coefficients.

In some problems a system which becomes active at $t = 0$, because of some initial disturbance, is subsequently acted upon by another disturbance beginning at a later time, say $t = a$. The analytical representation of such functions and the nature of their Laplace transforms are therefore a matter of some importance. To illustrate, suppose that we wish an expression describing the function whose graph is shown in Fig. 7.4a, the curve being congruent to the right half of the parabola $y = t^2$ shown in Fig. 7.4b. It is not enough to recall the translation formula from analytic geometry and write $f(t) = (t - a)^2$, because this equation, even with the usual qualification that $f(t) \equiv 0$ for $t < 0$, defines the curve shown in Fig. 7.4c and not the required graph. However, if we take the unit step function and translate it a units to the right by writing $u(t - a)$, we obtain the function shown in Fig. 7.4d. Since this vanishes for $t < a$ and is equal to 1 for $t > a$, the product $(t - a)^2 u(t - a)$ will be identically zero for $t < a$ and will be identically equal to $(t - a)^2$ for $t > a$ and hence will define precisely the arc we want. More generally, the expression

$$f(t - a)u(t - a)$$

represents the function obtained by translating $f(t)$ a units to the right and "cutting it off," i.e., making it vanish identically to the left of a.



(a) (b) (c) (d)

FIGURE 7.4

Plot describing the graph of a function which has been translated and "cut off."

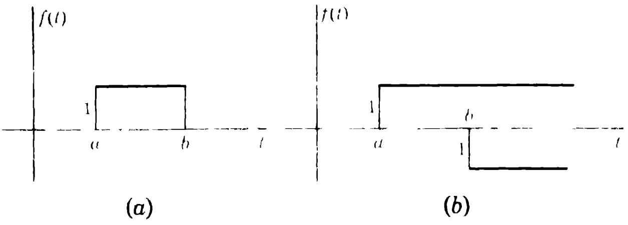
EXAMPLE 5

What is the equation of the function whose graph is shown in Fig. 7.5a?

Clearly we can regard this function as the sum of the two translated step functions shown in Fig. 7.5b. Hence its equation is

$$u(t - a) - u(t - b)$$

FIGURE 7.5
Plot showing how two step functions can be combined to give a rectangular pulse, or "filter function."

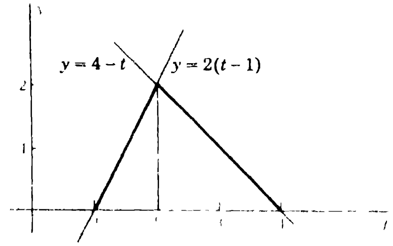


Although the function shown in Fig 7 5a is not ordinarily given a name, it could appropriately be referred to as a **filter function**. For when any other function is multiplied by this "filter function" it is annihilated completely, i e , reduced identically to zero, outside the "pass band," $a < t < b$, and reproduced without any change whatsoever for values of t within the "pass band "

EXAMPLE 6

What is the equation of the function whose graph is shown in Fig 7 6?

FIGURE 7.6
A graph consisting of straight-line segments



To obtain the segment of this function between 1 and 2 we must multiply the expression $2(t-1)$ by a factor which will be zero to the left of 1, unity between 1 and 2, and zero to the right of 2 By Example 5, such a function is $u(t-1) - u(t-2)$ Hence

$$2(t-1)[u(t-1) - u(t-2)]$$

defines the given function between 1 and 2 and vanishes elsewhere Similarly

$$(-t+4)[u(t-2) - u(t-4)]$$

defines the given function between 2 and 4 and vanishes elsewhere The complete representation of the function is therefore

$$\begin{aligned} &2(t-1)[u(t-1) - u(t-2)] + (-t+4)[u(t-2) - u(t-4)] \\ &\qquad = 2(t-1)u(t-1) - 3(t-2)u(t-2) + (t-4)u(t-4) \end{aligned}$$

The transforms of functions that have been translated and cut off are given by the so-called **second shifting theorem**:

THEOREM 6

$$\mathcal{L}\{f(t-a)u(t-a)\} = e^{-as}\mathcal{L}\{f(t)\} \qquad a \geqq 0$$

PROOF To prove this, we have by definition

$$\mathcal{L}\{f(t-a)u(t-a)\} = \int_0^\infty f(t-a)u(t-a)e^{-st}\,dt = \int_a^\infty f(t-a)e^{-st}\,dt$$

since the integration effectively commences not at $t = 0$ but at $t = a$ because $f(t-a)u(t-a)$ vanishes identically to the left of this point Now let $t - a = T$,

$dt = dT$. Then the last integral becomes

$$\int_0^\infty f(T)e^{-s(T+a)}\,dT = e^{-as}\int_0^\infty f(T)e^{-sT}\,dT = e^{-as}\mathcal{L}\{f(t)\} \qquad \text{as asserted.}$$

Before Theorem 6 can be applied, it is necessary that the function being transformed be expressed in terms of the binomial argument $t - a$ which appears in the unit step function. This will not often be the case, so it will frequently be necessary to alter the form of the function, as originally given, before it can be transformed. In many cases this can be done by inspection. On the other hand, we can always proceed in the following general way. Suppose we wish to transform

$$f(t)u(t-a)$$

As it stands, this cannot be handled by Theorem 6, so we rewrite it in the form

$$f[(t-a)+a]u(t-a) \equiv F(t-a)u(t-a)$$

where $F(t-a) = f[(t-a)+a] = f(t)$, or $F(t) = f(t+a)$. Now Theorem 6 can be applied, and we have

$$\mathcal{L}\{f(t)u(t-a)\} = \mathcal{L}\{F(t-a)u(t-a)\} = e^{-as}\mathcal{L}\{F(t)\} = e^{-as}\mathcal{L}\{f(t+a)\}$$

Thus we have established the following useful result.

COROLLARY 1

$$\mathcal{L}\{f(t)u(t-a)\} = e^{-as}\mathcal{L}\{f(t+a)\}$$

As a tool for finding inverses, it is convenient to restate Theorem 6 in the following form.

COROLLARY 2

If $\mathcal{L}^{-1}\{\phi(s)\} = f(t)$, then $\mathcal{L}^{-1}\{e^{-as}\phi(s)\} = f(t-a)u(t-a)$

In words, this says that *suppressing the factor e^{-as} in a transform requires that the inverse of what remains be translated a units to the right and cut off to the left of the point $t = a$.*

EXAMPLE 7

What is the transform of the function whose graph is shown in Fig. 7.7?

The equation of this function is obviously

$$\begin{aligned} F(t) &= -(t^2 - 3t + 2)[u(t-1) - u(t-2)] \\ &= -f(t)u(t-1) + f(t)u(t-2) \end{aligned}$$

where $f(t) = t^2 - 3t + 2$. However, the form of $f(t)$ is such that Theorem 6 cannot be applied

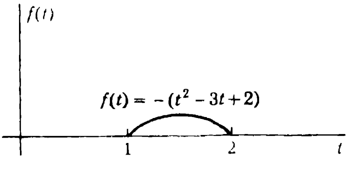


FIGURE 7.7
A parabolic pulse

directly to either term in the expression for $F(t)$. Hence we use Corollary 1, observing that

$$f(t+1) = [(t+1)^2 - 3(t+1) + 2] = t^2 - t$$

and

$$f(t+2) = [(t+2)^2 - 3(t+2) + 2] = t^2 + t$$

The required transform is, therefore,

$$-e^{-s}\mathcal{L}\{t^2 - t\} + e^{-2s}\mathcal{L}\{t^2 + t\} = -e^{-s}\left(\frac{2}{s^3} - \frac{1}{s^2}\right) + e^{-2s}\left(\frac{2}{s^3} + \frac{1}{s^2}\right)$$

EXAMPLE 8

Find the solution of the equation $y' + 3y + 2\int_0^t y\,dt = f(t)$ for which $y_0 = 1$, if $f(t)$ is the function whose graph is shown in Fig. 7.8.

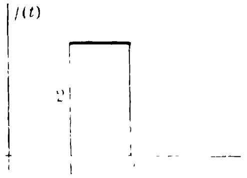


FIGURE 7.8 A rectangular pulse

In this case $f(t) = 2u(t-1) - 2u(t-2)$, and thus the differential equation can be written

$$y' + 3y + 2\int_0^t y\,dt = 2u(t-1) - 2u(t-2)$$

Taking transforms, we have

$$(s\mathcal{L}\{y\} - 1) + 3\mathcal{L}\{y\} + \frac{2}{s}\mathcal{L}\{y\} = \frac{2e^{-s}}{s} - \frac{2e^{-2s}}{s}$$

or

$$(s^2 + 3s + 2)\mathcal{L}\{y\} = 2e^{-s} - 2e^{-2s} + s$$

and

$$\mathcal{L}\{y\} = \frac{s}{(s+1)(s+2)} + \frac{2e^{-s}}{(s+1)(s+2)} - \frac{2e^{-2s}}{(s+1)(s+2)}$$

The first term can be written

$$\frac{2}{s+2} - \frac{1}{s+1}$$

Hence its inverse is $2e^{-2t} - e^{-t}$. If the exponential factors are suppressed in the second and third terms of $\mathcal{L}\{y\}$, the algebraic portion which remains can be written

$$2\left(\frac{1}{s+1} - \frac{1}{s+2}\right)$$

and the inverse of this is $2e^{-t} - 2e^{-2t}$. However, because the factors e^{-s} and e^{-2s} were neglected, it is necessary to take the last expression, translate it one unit to the right and cut it off to the left of $t = 1$, and also translate it two units to the right and cut it off to the left of $t = 2$ in order to obtain the inverses of the original terms. This gives for y

$$y = (2e^{-2t} - e^{-t}) + 2(e^{-(t-1)} - e^{-2(t-1)})u(t-1) - 2(e^{-(t-2)} - e^{-2(t-2)})u(t-2)$$

Plots of these three terms, as well as of their sum, that is, y itself, are shown in Fig. 7.9.

We have already made repeated use of Theorems 2 and 3 of Sec. 7.2 on the transforms of derivatives and integrals. On the other hand, it is sometimes convenient or necessary to consider the derivatives and integrals of transforms. The basis for this is contained in the next two theorems.

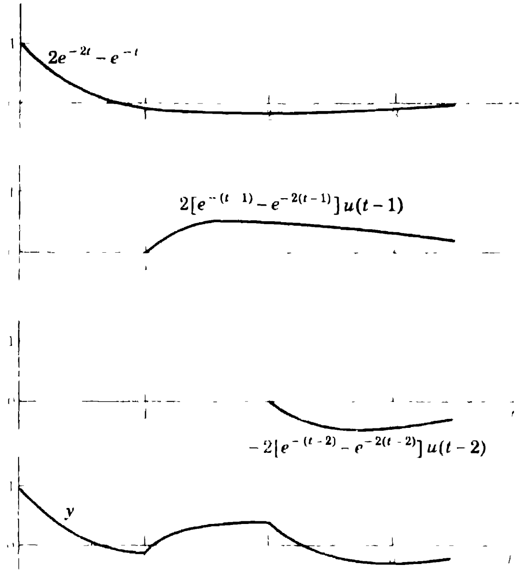


FIGURE 7.9
Plot showing the solution of Example 8

THEOREM 7
If $f(t)$ is piecewise regular and of exponential order and if $\mathcal{L}\{f(t)\} = \phi(s)$, then $\mathcal{L}\{tf(t)\} = -\phi'(s)$

PROOF By definition we have

$$\mathcal{L}\{f(t)\} = \int_0^\infty f(t)e^{-st}\,dt = \phi(s)$$

and, differentiating this with respect to s, we obtain

$$\frac{d}{ds}\int_0^\infty f(t)e^{-st}\,dt = \phi'(s)$$

Now under our usual assumptions that $f(t)$ is piecewise regular and of exponential order, the product $tf(t)$ also satisfies these conditions Hence, by Theorem 6, Sec 7.1, the integral which results when $\mathcal{L}\{f(t)\}$ is differentiated partially with respect to s, namely,

$$\int_0^\infty -tf(t)e^{-st}\,dt$$

converges uniformly Therefore, according to Theorem 3, Sec 7.1, the integral for $\mathcal{L}\{f(t)\}$ can legitimately be differentiated with respect to s inside the integral sign Thus, performing the differentiation, we have

$$\phi'(s) = \int_0^\infty f(t)[-te^{-st}]\,dt$$

or

$$\int_0^\infty [tf(t)]e^{-st}\,dt \equiv \mathcal{L}\{tf(t)\} = -\phi'(s) \qquad \text{as asserted.}$$

By taking inverses in the assertion of Theorem 7 and then solving for $f(t)$, we obtain the following useful result:

COROLLARY 1

If $\mathcal{L}\{f(t)\} = \phi(s)$, then $f(t) \equiv \mathcal{L}^{-1}\{\phi(s)\} = -(1/t)\mathcal{L}^{-1}\{\phi'(s)\}$.

This is often helpful when the inverse of a transform cannot conveniently be found but the inverse of the derivative of the transform is known The extension of Theorem 7 and its corollary to repeated differentiation of transforms is obvious.

THEOREM 8

If $f(t)$ is piecewise regular and of exponential order, if $\mathcal{L}\{f(t)\} = \phi(s)$, and if $f(t)/t$ has a limit as t approaches zero from the right, then

$$\mathcal{L}\left\{\frac{f(t)}{t}\right\} = \int_s^\infty \phi(s)\,ds$$

PROOF By definition

$$\phi(s) = \mathcal{L}\{f(t)\} = \int_0^\infty f(t)e^{-st}\,dt$$

Hence, integrating from s to ∞, we obtain

$$\int_s^\infty \phi(s)\,ds = \int_s^\infty \left[\int_0^\infty f(t)e^{-st}\,dt\right] ds$$

Now under the assumption that $\lim\limits_{t\to 0^+} \dfrac{f(t)}{t}$ exists and that $f(t)$ itself is piecewise regular and of exponential order, it follows from Theorems 6 and 2, Sec. 7 1, that the integration with respect to s can be performed inside the integral sign, i e., that the order of integration in the repeated integral can be reversed Hence, performing the integration,

$$\begin{aligned}\int_s^\infty \phi(s)\,ds &= \int_0^\infty \int_s^\infty f(t)e^{-st}\,ds\,dt = \int_0^\infty f(t)\left[\frac{e^{-st}}{-t}\right]_s^\infty dt \\ &= \int_0^\infty \left[\frac{f(t)}{t}\right] e^{-st}\,dt \\ &= \mathcal{L}\left\{\frac{f(t)}{t}\right\} \qquad \text{as asserted}\end{aligned}$$

By taking inverses in the assertion of Theorem 8 and then solving for $f(t)$, we obtain the following result

COROLLARY 1

If $\mathcal{L}\{f(t)\} = \phi(s)$, then $f(t) \equiv \mathcal{L}^{-1}\{\phi(s)\} = t\mathcal{L}^{-1}\left\{\int_s^\infty \phi(s)\,ds\right\}$

This is often useful in finding inverses when the integral of a transform is simpler to work with than the transform itself The extension of Theorem 8 and its corollary to repeated integration of transforms is immediate

EXAMPLE 9

What is $\mathcal{L}\{t^2 \sin 2t\}$?

By a repeated application of Theorem 7, we have

$$\mathcal{L}\{t^2 \sin 2t\} = (-1)^2\frac{d^2\mathcal{L}\{\sin 2t\}}{ds^2} = \frac{d^2}{ds^2}\left(\frac{2}{s^2+4}\right) = \frac{12s^2-16}{(s^2+4)^3}$$

EXAMPLE 10

What is y if $\mathcal{L}\{y\} = \ln[(s+1)/(s-1)]$?

Using Corollary 1 of Theorem 7, we have immediately

$$y = -\frac{1}{t}\mathcal{L}^{-1}\left\{\frac{d}{ds}\left(\ln\frac{s+1}{s-1}\right)\right\} = -\frac{1}{t}\mathcal{L}^{-1}\left\{\frac{1}{s+1}-\frac{1}{s-1}\right\} = \frac{e^{-t}-e^{t}}{-t} = \frac{2\sinh t}{t}$$

EXAMPLE 11

What is $\mathcal{L}\{(\sin kt)/t\}$?

By Theorem 8, we have

$$\mathcal{L}\left\{\frac{\sin kt}{t}\right\} = \int_s^\infty \mathcal{L}\{\sin kt\}\,ds = \int_s^\infty \frac{k}{s^2+k^2}\,ds = \operatorname{Tan}^{-1}\frac{s}{k}\Big|_s^\infty$$

$$= \frac{\pi}{2} - \operatorname{Tan}^{-1}\frac{s}{k} = \operatorname{Cot}^{-1}\frac{s}{k}$$

EXAMPLE 12

What is y if $\mathcal{L}\{y\} = s/(s^2-1)^2$?

Using Corollary 1 of Theorem 8, we have immediately

$$y = t\mathcal{L}^{-1}\left\{\int_s^\infty \frac{s}{(s^2-1)^2}\,ds\right\} = t\mathcal{L}^{-1}\left\{\frac{-1}{2(s^2-1)}\Big|_s^\infty\right\}$$

$$= t\mathcal{L}^{-1}\left\{\frac{1}{4}\left(\frac{1}{s-1}-\frac{1}{s+1}\right)\right\}$$

$$= \frac{t}{4}(e^{t}-e^{-t}) = \frac{t\sinh t}{2}$$

EXERCISES

Find the Laplace transform of each of the following functions

1 $u(t-a)$

2 $\cos(t-1)u(t-1)$

3 $t^2u(t-2)$

4 $(t^2-1)u(t-1)$

5 $e^{2t}u(t-1)$

6 $\cos 3t\, u(t-3)$

7 $f(t) = \begin{cases} \sin t & 0 < t < \pi \\ 0 & \pi < t \end{cases}$

8 $f(t) = \begin{cases} t & 0 < t < 2 \\ 2 & 2 < t \end{cases}$

9 See Fig 7 10

10 See Fig 7 11

FIGURE 7 10

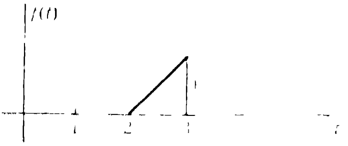
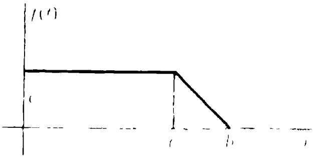


FIGURE 7.11



11 $\dfrac{1-\cos 3t}{t}$

12 $\dfrac{e^{2t}-1}{t}$

13 $te^{-3t}\sin 2t$

14 $t\int_0^t e^{-3t}\sin 2t\,dt$

15 $e^{-3t}\int_0^t t\sin 2t\,dt$

16 $\int_0^t te^{-3t}\sin 2t\,dt$

17 $\dfrac{e^{-3t}\sin 2t}{t}$

18 $e^{-3t}\int_0^t \dfrac{\sin 2t}{t}\,dt$

19 $\int_0^t \dfrac{e^{-3t}\sin 2t}{t}\,dt$

20 $\int_0^t \dfrac{e^t-\cos 2t}{t}\,dt$

Find the inverse of each of the following transforms:

21 $\dfrac{1}{(s+2)^4}$

22 $\dfrac{s}{(s+2)^4}$

23 $\dfrac{s+1}{9s^2+6s+5}$

24 $\dfrac{1}{s(s+2)^2}$

25 $\dfrac{1}{s^2(s+1)}$

26 $\dfrac{1}{(s+1)(s^2+2s+5)}$

27 $\dfrac{e^{-2s}}{s^2+4}$

28 $\dfrac{e^{-3s}}{s^2-9}$

29 $\dfrac{e^{-s}}{(s+1)^3}$

30 $\dfrac{e^{-s}+e^{-2s}}{(s-1)(s-2)}$

31 $\ln\dfrac{s+a}{s+b}$

32 $\ln\dfrac{s^2-1}{s^2}$

33 $\ln\dfrac{s^2+1}{s(s+1)}$

34 $s\ln\dfrac{s-1}{s+1}+2$

35 $\dfrac{1}{s}\operatorname{Tan}^{-1}\dfrac{1}{s}$

36 $\operatorname{Tanh}^{-1}\dfrac{1}{s}$

37 $\dfrac{s+2}{(s^2+4s+5)^2}$

38 $\dfrac{2s+3}{(s^2+3s+2)^2}$

39 $\dfrac{2}{(s^2+1)^2}$ (Hint. Multiply and divide the transform by s.)

40 Find the values of $f(0^+)$ and of $\lim\limits_{t\to\infty} f(t)$, if it exists, if $\mathcal{L}\{f(t)\}$ is

a $\dfrac{s^2+1}{s^3+6s^2+11s+6}$

b $\dfrac{s+3}{2s^3-3s^2-2s}$

c $\dfrac{s^2+s+1}{s^3-s^2+2}$

d $\dfrac{s+2}{s^3+3s^2+4s+2}$

41 Show that, under appropriate conditions,

$$\lim_{s\to\infty} s[s\mathcal{L}\{f(t)\}-f(0^+)] = f'(0^+)$$

and that

$$\lim_{s\to\infty} s[s^2\mathcal{L}\{f(t)\}-sf(0^+)-f'(0^+)] = f''(0^+)$$

What conditions beyond those of Theorem 1 are necessary for the validity of these results? Can the value of $f^{(n)}(0^+)$ be obtained by an extension of these formulas?

42 Show that, under appropriate conditions,

$$\lim_{s\to 0} s[s\mathcal{L}\{f(t)\}-f(0^+)] = \lim_{t\to\infty} f'(t)$$

What conditions beyond those of Theorem 2 are necessary for the validity of this result? Can this result be generalized to the determination of $\lim\limits_{t\to\infty} f^{(n)}(t)$ from $\mathcal{L}\{f(t)\}$?

Solve the following differential equations.

43 $y'' + 4y' + 3y = e^{-t}$ $\qquad y_0 = y_0' = 1$

44 $y'' + 4y = \cos 2t$ $\qquad y_0 = -2,\ y_0' = 1$

45 $y'' + 3y' + 2y = u(t-1)$ $\qquad y_0 = 0,\ y_0' = 1$

46 $y'' + 4y' + 4y = (t-2)e^{-(t-2)}u(t-2)$ $\qquad y_0 = 1,\ y_0' = -1$

47 $y^{IV} + 2y'' + y = 0$ $\qquad y_0 = y_0' = y_0''' = 0,\ y_0'' = 1$

48 Prove Theorem 1 without assuming that $f(t)$ is continuous. (Hint: Use the result of Exercise 3, Sec. 7.2.)

49 Prove Theorem 2 without assuming that $f(t)$ is continuous. (Hint: Use the result of Exercise 3, Sec. 7.2.)

50 Where in the proof of Theorem 2 is use made of the hypothesis that the abscissa of convergence of $f(t)$ is negative?

7.5

The Heaviside expansion theorems

The frequent use we have had to make of partial fractions indicates clearly the importance of this technique in operational calculus. It is therefore highly desirable to have the procedure systematized as much as possible. The following theorems, usually associated with the name of Heaviside, are of great utility in this connection:

THEOREM 1

If $f(t) = \mathcal{L}^{-1}\{p(s)/q(s)\}$, where $p(s)$ and $q(s)$ are polynomials and the degree of $q(s)$ is greater than the degree of $p(s)$, then the term in $f(t)$ corresponding to an unrepeated linear factor $s - a$ of $q(s)$ is

$$\frac{p(a)}{q'(a)}e^{at} \quad \text{or equally well} \quad \frac{p(a)}{Q(a)}e^{at}$$

where $Q(s)$ is the product of all the factors of $q(s)$ except $s - a$.

PROOF In the familiar partial-fraction decomposition of $p(s)/q(s)$, an unrepeated linear factor $s - a$ of $q(s)$ gives rise to a single fraction of the form $A/(s-a)$. Hence, if we denote by $h(s)$ the sum of the fractions corresponding to all the other factors of $q(s)$, we can write

$$\frac{p(s)}{q(s)} = \frac{A}{s-a} + h(s)$$

where, since $s - a$ is an unrepeated factor of $q(s)$, $h(s)$ remains finite as s approaches a. Multiplying this identity by $s - a$ then gives

$$\frac{(s-a)p(s)}{q(s)} \equiv \frac{p(s)}{q(s)/(s-a)} = A + (s-a)h(s)$$

If we now let s approach a, the second term in the right member vanishes, and we have

$$A = \lim_{s\to a}\frac{p(s)}{q(s)/(s-a)}$$

The limit of the numerator here is evidently $p(a)$. The denominator appears as an indeterminate of the form $0/0$. However, if we evaluate it as usual according to L'Hospital's rule by differentiating numerator and denominator and then letting s approach a, we obtain just $q'(a)$. Hence

$$A = \frac{p(a)}{q'(a)}$$

On the other hand, we could have eliminated the indeterminacy before passing to the limit simply by canceling $s - a$ into $q(s)$, which by hypothesis contains this factor Doing this, we obtain the equivalent form of A.

$$A = \frac{p(a)}{Q(a)}$$

Finally, taking inverses, it is clear that the fraction $A/(s - a)$ gives rise to the term

$$Ae^{at} = \frac{p(a)}{q'(a)} e^{at} = \frac{p(a)}{Q(a)} e^{at}$$

in the inverse $f(t)$, as asserted.

If $q(s)$ contains only unrepeated linear factors, then by applying Theorem 1 to each factor in turn, we obtain the following useful result

COROLLARY 1

If $f(t) = \mathcal{L}^{-1}\{p(s)/q(s)\}$ and if $q(s)$ is completely factorable into unrepeated linear factors

$$(s - a_1), \qquad (s - a_2), \qquad \ldots, \qquad (s - a_n)$$

then

$$f(t) = \sum_{i=1}^{n} \frac{p(a_i)}{q'(a_i)} e^{a_i t} = \sum_{i=1}^{n} \frac{p(a_i)}{Q_i(a_i)} e^{a_i t}$$

where $Q_i(s)$ is the product of all the factors of $q(s)$ except the factor $s - a_i$

THEOREM 2

If $f(t) = \mathcal{L}^{-1}\{p(s)/q(s)\}$, where $p(s)$ and $q(s)$ are polynomials and the degree of $q(s)$ is greater than the degree of $p(s)$, then the terms in $f(t)$ corresponding to a repeated linear factor $(s - a)^r$ in $q(s)$ are

$$\left[\frac{\phi^{(r-1)}(a)}{(r-1)!} + \frac{\phi^{(r-2)}(a)}{(r-2)!}\frac{t}{1!} + \cdots + \frac{\phi'(a)}{1!}\frac{t^{r-2}}{(r-2)!} + \phi(a)\frac{t^{r-1}}{(r-1)!}\right] e^{at}$$

where $\phi(s)$ is the quotient of $p(s)$ and all the factors of $q(s)$ except $(s - a)^r$

PROOF From the familiar theory of partial fractions we recall that a repeated linear factor $(s - a)^r$ of $q(s)$ gives rise to the component fractions

$$\frac{A_1}{s-a} + \frac{A_2}{(s-a)^2} + \cdots + \frac{A_{r-1}}{(s-a)^{r-1}} + \frac{A_r}{(s-a)^r}$$

If we let $h(s)$ denote, as before, the sum of the fractions corresponding to all the other factors of $q(s)$, we have

$$\frac{p(s)}{q(s)} = \frac{\phi(s)}{(s-a)^r} = \frac{A_1}{s-a} + \frac{A_2}{(s-a)^2} + \cdots + \frac{A_{r-1}}{(s-a)^{r-1}} + \frac{A_r}{(s-a)^r} + h(s)$$

Multiplying this identity by $(s-a)^r$ gives

$$\phi(s) = A_1(s-a)^{r-1} + A_2(s-a)^{r-2} + \cdots + A_{r-1}(s-a) + A_r + (s-a)^r h(s)$$

If we put $s = a$ in this expression, we obtain

$$\phi(a) = A_r$$

If we now differentiate $\phi(s)$, we have

$$\begin{aligned}\phi'(s) = A_1(r-1)(s-a)^{r-2} + A_2(r-2)(s-a)^{r-3} + \cdots \\ + A_{r-1} + r(s-a)^{r-1}h(s) + (s-a)^r h'(s)\end{aligned}$$

Again setting $s = a$, we find this time

$$\phi'(a) = A_{r-1}$$

Continuing in this fashion, noting that the first $r-1$ derivatives of the product $(s-a)^r h(s)$ will all vanish when $s = a$, we obtain successively

$$\begin{aligned}\phi''(a) &= 2!A_{r-2} \\ \phi'''(a) &= 3!A_{r-3} \\ &\cdots \\ \phi^{(r-1)}(a) &= (r-1)!A_1\end{aligned}$$

or

$$A_{r-k} = \frac{\phi^{(k)}(a)}{k!} \qquad k = 0, 1, \ldots, r-1$$

The terms in the expansion of $p(s)/q(s)$ which correspond to the factor $(s-a)^r$ are, therefore,

$$\begin{aligned}\frac{\phi^{(r-1)}(a)}{(r-1)!}\,\frac{1}{s-a} + \frac{\phi^{(r-2)}(a)}{(r-2)!}\cdot\frac{1}{(s-a)^2} + \cdots \\ + \frac{\phi'(a)}{1!}\,\frac{1}{(s-a)^{r-1}} + \phi(a)\,\frac{1}{(s-a)^r}\end{aligned}$$

Recalling that

$$\mathcal{L}^{-1}\left\{\frac{1}{(s-a)^n}\right\} = \frac{t^{n-1}e^{at}}{(n-1)!}$$

it is evident that the terms in y which arise from these fractions are

$$\frac{\phi^{(r-1)}(a)}{(r-1)!}e^{at} + \frac{\phi^{(r-2)}(a)}{(r-2)!}\,\frac{te^{at}}{1!} + \cdots + \frac{\phi'(a)}{1!}\,\frac{t^{r-2}e^{at}}{(r-2)!} + \phi(a)\,\frac{t^{r-1}e^{at}}{(r-1)!}$$

If we factor out e^{at} from this expression, we have precisely the assertion of the theorem

THEOREM 3

If $f(t) = \mathcal{L}^{-1}\{p(s)/q(s)\}$, where $p(s)$ and $q(s)$ are polynomials and the degree of $q(s)$ is greater than the degree of $p(s)$, then the terms in $f(t)$ which correspond to an unrepeated, irreducible quadratic factor $(s+a)^2 + b^2$ of $q(s)$ are

$$\frac{e^{-at}}{b}\,(\phi_i \cos bt + \phi_r \sin bt)$$

where ϕ_r and ϕ_i are, respectively, the real and imaginary parts of $\phi(-a+ib)$, and $\phi(s)$ is the quotient of $p(s)$ and all the factors of $q(s)$ except the factor $(s+a)^2 + b^2$.

PROOF From the familiar theory of partial fractions we recall that an unrepeated, irreducible quadratic factor $(s + a)^2 + b^2$ of $q(s)$ gives rise to a single fraction of the form

$$\frac{As + B}{(s + a)^2 + b^2}$$

in the partial-fraction expansion of $p(s)/q(s)$ If again we let $h(s)$ denote the fractions corresponding to all the other factors of $q(s)$, we can, therefore, write

$$\frac{p(s)}{q(s)} \equiv \frac{\phi(s)}{(s + a)^2 + b^2} = \frac{As + B}{(s + a)^2 + b^2} + h(s)$$

Multiplying this identity by $(s + a)^2 + b^2$, we obtain

$$\phi(s) = As + B + [(s + a)^2 + b^2]h(s)$$

Now put $s = -a + ib$ This value, of course, makes $(s + a)^2 + b^2$ vanish, hence the last product drops out, leaving

$$\phi(-a + ib) = (-a + ib)A + B$$

or, reducing $\phi(-a + ib)$ to its standard complex form $\phi_r + i\phi_i$,

$$\phi_r + i\phi_i = (-aA + B) + ibA$$

Equating real and imaginary terms in the last identity, we find

$$\phi_r = -aA + B \qquad \phi_i = bA$$

or, solving for A and B,

$$A = \frac{\phi_i}{b} \qquad B = \frac{b\phi_r + a\phi_i}{b}$$

Thus the partial fraction which corresponds to the quadratic factor $(s + a)^2 + b^2$ is

$$\begin{aligned} \frac{As + B}{(s + a)^2 + b^2} &= \frac{1}{b}\,\frac{\phi_i s + (b\phi_r + a\phi_i)}{(s + a)^2 + b^2} \\ &= \frac{1}{b}\left[\frac{(s + a)\phi_i}{(s + a)^2 + b^2} + \frac{b\phi_r}{(s + a)^2 + b^2}\right] \end{aligned}$$

The inverse of this expression is evidently

$$\frac{1}{b}\,(\phi_i e^{-at} \cos bt + \phi_r e^{-at} \sin bt)$$

Factoring out e^{-at} now gives the assertion of the theorem

There is a fourth theorem dealing with repeated, irreducible quadratic factors, but, because of its complexity and limited usefulness, we shall not develop it here Fortunately, many of the simpler transforms involving repeated quadratic factors can be handled by other means, for instance, the convolution theorem of Sec. 7 7.

EXAMPLE 1

If $\mathcal{L}\{f\} = (s^2 + 2)/s(s + 1)(s + 2)$, what is $f(t)$?

The roots of the denominator are $s = 0, -1, -2$ Hence we must compute the values of

$$p(s) = s^2 + 2 \qquad \text{and} \qquad q'(s) = 3s^2 + 6s + 2$$

for these values of s The results are

$$p(0) = 2 \qquad p(-1) = 3 \qquad p(-2) = 6$$
$$q'(0) = 2 \qquad q'(-1) = -1 \qquad q'(-2) = 2$$

From the corollary of Theorem 1 we now have at once

$$f(t) = \frac{2}{2}e^{0t} + \frac{3}{-1}e^{-t} + \frac{6}{2}e^{-2t} = 1 - 3e^{-t} + 3e^{-2t}$$

Equally well, of course, we could have obtained the coefficients in the inverse by suppressing each of the factors in turn and evaluating the rest of the fraction at the root associated with the suppressed factor

EXAMPLE 2

If $\mathcal{L}\{y\} = s/(s + 2)^2(s^2 + 2s + 10)$, what is y?

Considering first the repeated linear factor, we identify

$$\phi(s) = \frac{s}{s^2 + 2s + 10} \qquad \text{and} \qquad \phi'(s) = \frac{-s^2 + 10}{(s^2 + 2s + 10)^2}$$

Evaluating these for the root $s = -2$, we obtain

$$\phi(-2) = -\tfrac{1}{5} \qquad \text{and} \qquad \phi'(-2) = \tfrac{3}{50}$$

Hence, by Theorem 2, the terms in y corresponding to $(s + 2)^2$ are

$$e^{-2t}\left(\frac{3}{50} - \frac{t}{5}\right) = \frac{(3 - 10t)e^{-2t}}{50}$$

For the quadratic factor $s^2 + 2s + 10 \equiv (s + 1)^2 + 3^2$, we have

$$\phi(s) = \frac{s}{(s + 2)^2}$$

Hence,

$$\phi(-a + ib) = \phi(-1 + 3i) = \frac{-1 + 3i}{[(-1 + 3i) + 2]^2} = \frac{-1 + 3i}{(1 + 3i)^2} = \frac{-1 + 3i}{-8 + 6i} = \frac{13 - 9i}{50}$$

and thus $\phi_r = {}^{13}\!/_{50}$, $\phi_i = -{}^{9}\!/_{50}$ The term in y corresponding to the factor

$$s^2 + 2s + 10$$

is, therefore,

$$\frac{1}{3}\left[\frac{e^{-t}(-9 \cos 3t + 13 \sin 3t)}{50}\right]$$

Adding the two partial inverses, we have finally

$$y = \frac{(3 - 10t)e^{-2t}}{50} + \frac{e^{-t}(-9 \cos 3t + 13 \sin 3t)}{150}$$

EXERCISES

Find the functions which have the following transforms

1 $\dfrac{s^2 - s + 3}{s^3 + 6s^2 + 11s + 6}$

2 $\dfrac{s + 2}{(s + 1)(s^2 + 4)}$

3 $\dfrac{s}{(s + 2)^2(s^2 + 1)}$

4 $\dfrac{s + 1}{(s^2 + 1)(s^2 + 4s + 13)}$

5 $\dfrac{s + 2}{s^4 + 4s^3 + 4s^2 - 4s - 5}$

6 $\dfrac{s}{(s + 1)(s + 2)^3}$

7 $\dfrac{s}{s^4 - 2s^2 + 1}$

8 $\dfrac{s + 2}{s^4 - 16s^2 + 100}$

Solve the following differential equations:

9 $y''' - 2y'' - y' + 2y = u(t-2) \qquad y_0 = y_0' = 0,\ y_0'' = 1$

10 $y''' + 3y'' + 3y' + y = \cosh t \qquad y_0 = y_0' = y_0'' = 0$

11 $y^{\text{iv}} + 2y''' + 2y'' + 2y' + y = e^{-t} \qquad y_0 = y_0' = y_0'' = y_0''' = 0$

12 $\left.\begin{aligned} x'' + 2x' + \int_0^t y\,dt &= t \\ 4x'' - 5x' + y &= \sin 2t \end{aligned}\right\} \qquad x_0 = -1,\ x_0' = 1$

13 $\left.\begin{aligned} (D^2 + D + 1)x + (D-1)y &= u(t) \\ (D^2 + 2D + 3)x + (3D^2 + 4D - 3)y &= u(t-1) \end{aligned}\right\} \qquad x_0 = x_0' = y_0 = y_0' = 0$

14 $\left.\begin{aligned} y' - 3z &= 5 \\ y + z' - w &= 3 - 2t \\ z + w' &= -1 \end{aligned}\right\} \qquad y_0 = 1,\ z_0 = 0,\ w_0 = -1$

15 In the proof of Theorem 3, verify that, if the identity

$$\phi(s) = As + B + [(s+a)^2 + b^2]h(s)$$

is evaluated for $s = -a - ib$ instead of for $s = -a + ib$, the same inverse is obtained

7.6

Transforms of periodic functions

The application of the Laplace transformation to the important case of general periodic functions is based upon the following theorem

THEOREM 1

If $f(t)$ is a piecewise regular function of exponential order which is periodic with period a, then

$$\mathcal{L}\{f(t)\} = \frac{\int_0^a f(t)e^{-st}\,dt}{1 - e^{-as}}$$

PROOF By definition,

$$\begin{aligned} \mathcal{L}\{f(t)\} &= \int_0^\infty f(t)e^{-st}\,dt \\ &= \int_0^a f(t)e^{-st}\,dt + \int_a^{2a} f(t)e^{-st}\,dt + \int_{2a}^{3a} f(t)e^{-st}\,dt + \cdots \end{aligned}$$

Now, in the second integral, let $t = T + a$, in the third integral let $t = T + 2a$; and in general let $t = T + na$ in the $(n+1)$st integral In each case $dt = dT$, and the new limits become 0 and a Hence

$$\begin{aligned} \mathcal{L}\{f(t)\} &= \int_0^a f(T)e^{-sT}\,dT + \int_0^a f(T+a)e^{-s(T+a)}\,dT \\ &\qquad + \int_0^a f(T+2a)e^{-s(T+2a)}\,dT + \cdots \\ &= \int_0^a f(T)e^{-sT}\,dT + e^{-as}\int_0^a f(T+a)e^{-sT}\,dT \\ &\qquad + e^{-2as}\int_0^a f(T+2a)e^{-sT}\,dT + \cdots \end{aligned}$$

But $f(T+a) = f(T+2a) = \cdots = f(T+na) = \cdots = f(T)$ for all values of T, since, by hypothesis, $f(t)$ is of period a. Thus we have

$$\mathcal{L}\{f(t)\} = \int_0^a f(T)e^{-sT}\,dT + e^{-as}\int_0^a f(T)e^{-sT}\,dT + e^{-2as}\int_0^a f(T)e^{-sT}\,dT + \cdots$$
$$= (1 + e^{-as} + e^{-2as} + \cdots)\int_0^a f(T)e^{-sT}\,dT$$

Now, if the infinite geometric progression which multiplies the integral is explicitly summed, using the familiar formula $S = 1/(1 - r)$, where the common ratio r is e^{-as}, we obtain the result of the theorem.

EXAMPLE 1

Find the transform of the rectangular wave shown in Fig. 7.12.

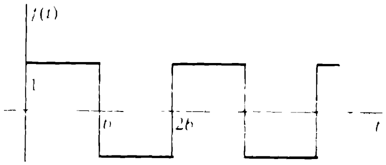


FIGURE 7.12
An alternating rectangular wave

The period here is $2b$. Hence by Theorem 1,

$$\mathcal{L}\{f(t)\} = \frac{1}{1-e^{-2bs}}\int_0^{2b} f(t)e^{-st}\,dt$$
$$= \frac{1}{1-e^{-2bs}}\left[\int_0^b 1\cdot e^{-st}\,dt + \int_b^{2b} -1\cdot e^{-st}\,dt\right]$$
$$= \frac{1}{1-e^{-2bs}}\cdot\frac{1-2e^{-bs}+e^{-2bs}}{s} = \frac{(1-e^{-bs})^2}{s(1-e^{-bs})(1+e^{-bs})}$$
$$= \frac{1-e^{-bs}}{s(1+e^{-bs})} = \frac{e^{bs/2}-e^{-bs/2}}{s(e^{bs/2}+e^{-bs/2})} = \frac{1}{s}\tanh\frac{bs}{2}$$

EXAMPLE 2

Find the transform of the saw-tooth wave shown in Fig. 7.13.

Here the period is k, and thus

$$\mathcal{L}\{f(t)\} = \frac{1}{1-e^{-ks}}\int_0^k te^{-st}\,dt = \frac{1}{1-e^{-ks}}\left[\frac{e^{-st}}{s^2}(-st-1)\right]_0^k$$
$$= \frac{1-(1+ks)e^{-ks}}{s^2(1-e^{-ks})} = \frac{(1+ks)-(1+ks)e^{-ks}-ks}{s^2(1-e^{-ks})}$$
$$= \frac{1+ks}{s^2} - \frac{k}{s(1-e^{-ks})}$$

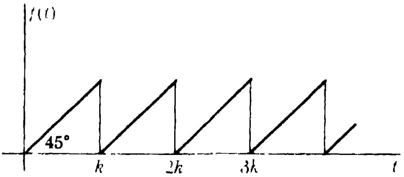


FIGURE 7.13
A saw-tooth wave

EXAMPLE 3

What is the Laplace transform of the **staircase function**

$$f(t) = n + 1 \qquad nk < t < (n+1)k \qquad n = 0, 1, 2, \ldots$$

shown in Fig 7 14a?

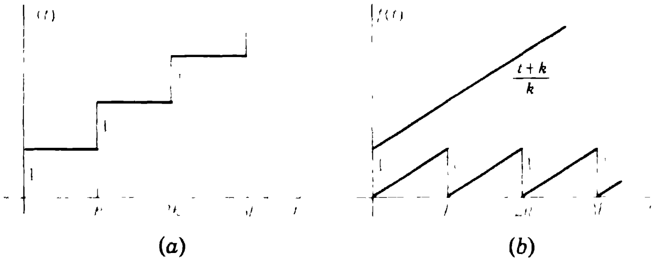


FIGURE 7.14 The "staircase function" and its synthesis.

The required transform can easily be found by direct calculation. However, it is even simpler to obtain it by considering $f(t)$ to be the difference of the two functions shown in Fig 7 14b The transform of the linear function $(t + k)/k$ can be found at once by Formula 4, Sec 7.3 Except for the obvious coefficient $1/k$, the transform of the saw-tooth function was obtained in the last example Hence,

$$\mathcal{L}\{f(t)\} = \frac{1}{k}\left(\frac{1}{s^2} + \frac{k}{s}\right) - \frac{1}{k}\left(\frac{1 + ks}{s^2} - \frac{k}{s(1 - e^{-ks})}\right) = \frac{1}{s(1 - e^{-ks})}$$

EXAMPLE 4

If the Laplace transform of $f(t)$ is $1/[(s + a)(1 - e^{-ks})]$, what is $f(t)$?

Although $\mathcal{L}\{f(t)\}$ resembles somewhat the transform of the staircase function obtained in the last example, the correspondence is not sufficiently close to provide us with the required inverse Moreover, we cannot successfully employ the result of the last example after first using the corollary of Theorem 5, Sec 7 4, for if we replace s by $s - a$, the given transform becomes

$$\frac{1}{s[1 - e^{-k(s-a)}]} = \frac{1}{s[1 - e^{ak}e^{-ks}]}$$

and now, because of the factor e^{ak}, which is not equal to 1 except in the trivial cases $a = 0$ or $k = 0$, we still do not have the transform of the staircase function It appears, therefore, that we must make a direct attack upon the problem To do this, let us reverse the derivation of Theorem 1 and replace $1/(1 - e^{-ks})$ by the infinite geometric series of which it is the sum

$$\mathcal{L}\{f(t)\} = \frac{1}{(s + a)(1 - e^{-ks})} = \frac{1}{s + a}(1 + e^{-ks} + e^{-2ks} + e^{-3ks} + \cdots)$$

$$= \frac{1}{s + a} + \frac{e^{-ks}}{s + a} + \frac{e^{-2ks}}{s + a} + \frac{e^{-3ks}}{s + a} + \cdots$$

Now let us assume that we can take the inverse of this infinite series term by term If we neglect the exponential in, say, the $(n + 1)$st term, the inverse of what remains is obvious, namely,

$$e^{-at}$$

But, having neglected the exponential e^{-nks}, we must, according to Corollary 2 of Theorem 6, Sec 7 4, translate the function e^{-at} to the right a distance of nk and then cut it off to the left of $t = nk$ When this is done for each term, we have

$$f(t) = e^{-at} + e^{-a(t-k)}u(t - k) + e^{-a(t-2k)}u(t - 2k) + e^{-a(t-3k)}u(t - 3k) + \cdots$$

Taking into account the "cutoff" properties of the various translated step functions, it is thus clear that the function $f(t)$ is equal to

$$\begin{array}{ll}
e^{-at} & \text{over the interval } (0,k) \\
e^{-at} + e^{ak}e^{-at} & \text{over the interval } (k,2k) \\
e^{-at} + e^{ak}e^{-at} + e^{2ak}e^{-at} & \text{over the interval } (2k,3k) \\
\cdots & \cdots \\
e^{-at} + e^{ak}e^{-at} + e^{2ak}e^{-at} + \cdots + e^{nak}e^{-at} & \text{over the interval } [nk,\ (n+1)k]
\end{array}$$

In order to obtain a more convenient expression for $f(t)$ over the general interval $nk < t < (n + 1)k$, we can sum the finite geometric progression defining $f(t)$ in this range. Since this progression contains $n + 1$ terms and has the common ratio $r = e^{ak}$, it follows that over this interval we have

$$f(t) = e^{-at}(1 + e^{ak} + e^{2ak} + \cdots + e^{nak}) = e^{-at}\frac{(e^{ak})^{n+1} - 1}{e^{ak} - 1}$$

$$= \frac{e^{-a[t-(n+1)k]}}{e^{ak} - 1} - \frac{e^{-at}}{e^{ak} - 1} \qquad nk < t < (n + 1)k$$

Now, to achieve a more symmetric form, let us define $\tau = t - (n + 1)k$. Clearly, $t = nk$ corresponds to $\tau = -k$ and $t = (n + 1)k$ corresponds to $\tau = 0$, so that, for each value of n, the parameter τ ranges from $-k$ to 0 as t ranges from nk to $(n + 1)k$. If we make this substitution in the first fraction only, $f(t)$ assumes the form

$$f(t) = \frac{e^{-a\tau}}{e^{ak} - 1} - \frac{e^{-at}}{e^{ak} - 1} \qquad -k < \tau < 0, \qquad nk < t < (n + 1)k$$

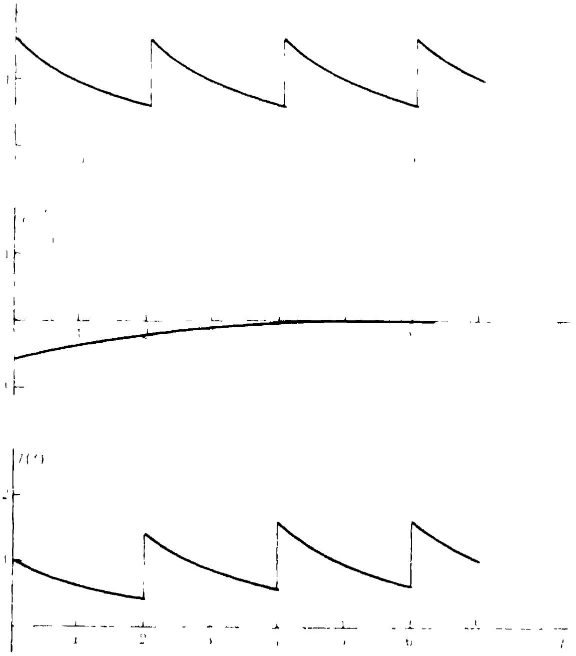


FIGURE 7.15
Plot showing the inverse of $\phi(s) = \dfrac{1}{(s + \frac{1}{2})(1 - e^{-2s})}$

The second term is a continuous function, dying away rapidly as t increases if $a > 0$. The first term is completely independent of n, that is, yields the same set of values over each interval, because no matter what n may be, as t ranges from nk to $(n + 1)k$, τ always ranges from $-k$ to 0. Moreover, the first term is discontinuous, since at the left end of any interval, where $\tau = -k$, its value is

$$\frac{e^{-a(-k)}}{e^{ak} - 1}$$

while at the right end, where $\tau = 0$, its value is

$$\frac{1}{e^{ak} - 1}$$

The periodic function it represents has, therefore, a jump of

$$\frac{e^{ak}}{e^{ak} - 1} - \frac{1}{e^{ak} - 1} = 1$$

at each of the points $t = k, 2k, 3k, \ldots$.

In Fig. 7.15 the discontinuous periodic function represented by the first term in $f(t)$, the continuous transient term represented by the second fraction, and $f(t)$ itself are shown for $a = \frac{1}{2}$ and $k = 2$.

EXAMPLE 5

What is the solution of the equation $y' + 3y + 2\int_0^t y\,dt = f(t)$ if $y_0 = 1$ and if $f(t)$ is the function shown in Fig. 7.16?

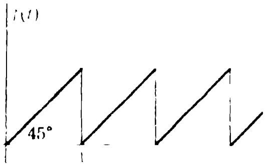


FIGURE 7.16 A saw-tooth wave

Taking the transform of each side of the given equation, using the result of Example 2 to transform $f(t)$, we have

$$(s\mathcal{L}\{y\} - 1) + 3\mathcal{L}\{y\} + \frac{2}{s}\mathcal{L}\{y\} = \frac{1 + s}{s^2} - \frac{1}{s(1 - e^{-s})}$$

or

$$\mathcal{L}\{y\} = \frac{s^2 + s + 1}{s(s + 1)(s + 2)} - \frac{1}{(s + 1)(s + 2)(1 - e^{-s})}$$

The inverse of the first fraction can be found immediately by the corollary of the first Heaviside theorem:

$$\tfrac{1}{2} - e^{-t} + \tfrac{3}{2}e^{-2t}$$

To find the inverse of the second fraction we must write

$$\frac{1}{(s + 1)(s + 2)(1 - e^{-s})} = \left(\frac{1}{s + 1} - \frac{1}{s + 2}\right)\frac{1}{1 - e^{-s}}$$

$$= \frac{1}{(s + 1)(1 - e^{-s})} - \frac{1}{(s + 2)(1 - e^{-s})}$$

and then use the results of Example 4. In this case $k = 1$, and thus the inverse over the general interval $n < t < n + 1$ is

$$\left(\frac{e^{-\tau}}{e-1} - \frac{e^{-t}}{e-1}\right) - \left(\frac{e^{-2\tau}}{e^2-1} - \frac{e^{-2t}}{e^2-1}\right)$$

or

$$\left(\frac{e^{-\tau}}{e-1} - \frac{e^{-2\tau}}{e^2-1}\right) - \left(\frac{e^{-t}}{e-1} - \frac{e^{-2t}}{e^2-1}\right) \qquad -1 < \tau < 0$$

The second term is obviously a continuous function of t and is simply an additional contribution to the transient of the system The periodic function defined by the first term is also continuous in this case, because the unit jumps exhibited by each of the fractions at $t = 1, 2, 3,$ are of opposite sign and, hence, cancel each other. The entire solution for y is therefore

$$y = \frac{1 - 2e^{-t} + 3e^{-2t}}{2} + \left(\frac{e^{-t}}{e-1} - \frac{e^{-2t}}{e^2-1}\right) - \left(\frac{e^{-\tau}}{e-1} - \frac{e^{-2\tau}}{e^2-1}\right)$$

$$= \underbrace{\left[-\frac{e-2}{e-1}e^{-t} + \frac{3e^2-5}{2(e^2-1)}e^{-2t}\right]}_{\text{transient}} + \underbrace{\left(\frac{1}{2} - \frac{e^{-\tau}}{e-1} + \frac{e^{-2\tau}}{e^2-1}\right)}_{\text{steady-state}} \qquad -1 < \tau < 0$$

Figure 7 17 shows a plot of the component terms and of y itself

The analysis of equations like the one considered in Example 5 is so important that a table of additional results similar to that obtained in Example 4 would be highly desirable. Using

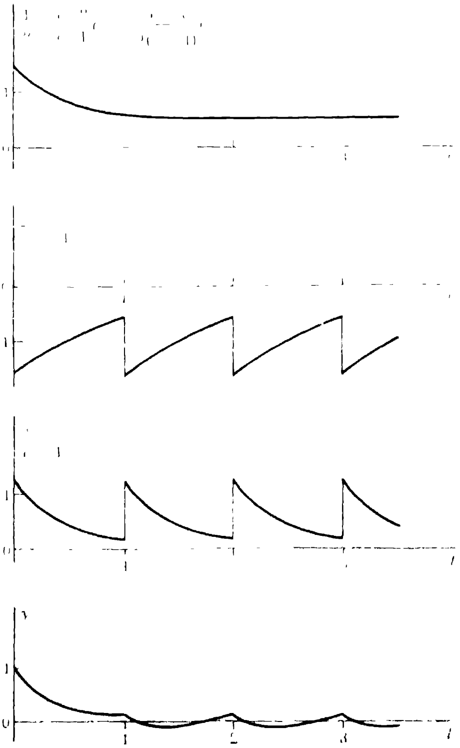


FIGURE 7.17
Plot showing the solution of Example 5

for the most part only the procedure illustrated in Example 4, such a table can easily be developed, as we shall now show

To eliminate unnecessary writing, it will be convenient to introduce the functions defined in Table 7 1 for the interval $nk < x < (n+1)k$, where k is an arbitrary positive number, n is an arbitrary nonnegative integer, and x is a variable which is to be replaced by t or τ, as required The functions $\phi_1(x,k)$ and $\phi_2(x,k)$ are, respectively, the staircase function and the **Morse dot function.** The functions $\phi_3(x,k)$ and $\phi_4(x,k)$ are the integrals from 0 to x of $\phi_1(t,k)$ and $\phi_2(t,k)$, respectively The function $\phi_5(x,a,k)$ is precisely that which we encountered in the solution of Example 4 The others, though somewhat more complicated,

table 7.1

Definition of functional symbol	Definition of function over general interval $nk < x < (n+1)k$
$\phi_1(x,k)$	$n+1$
$\phi_2(x,k)$	$\dfrac{(-1)^n+1}{2}$
$\phi_3(x,k)$	$(n+1)x - \dfrac{n(n+1)k}{2}$
$\phi_4(x,k)$	$\dfrac{(-1)^n+1}{2}x + \dfrac{k}{4}[1-(-1)^n(2n+1)]$
$\phi_5(x,a,k)$	$\dfrac{e^{-ax}}{e^{ak}-1}$
$\phi_6(x,a,k)$	$\dfrac{e^{-ax}}{e^{ak}+1}$
$\phi_7(x,a,b,k)$	$\dfrac{e^{-ax}\cos b(x+k) - e^{-a(x+k)}\cos bx}{2(\cosh ak - \cos bk)}$
$\phi_8(x,a,b,k)$	$\dfrac{e^{-ax}\cos b(x+k) + e^{-a(x+k)}\cos bx}{2(\cosh ak + \cos bk)}$
$\phi_9(x,a,b,k)$	$\dfrac{e^{-ax}\sin b(x+k) - e^{-a(x+k)}\sin bx}{2(\cosh ak - \cos bk)}$
$\phi_{10}(x,a,b,k)$	$\dfrac{e^{-ax}\sin b(x+k) + e^{-a(x+k)}\sin bx}{2(\cosh ak + \cos bk)}$
$\phi_{11}(x,a,k)$	$\dfrac{(x+k)e^{-ax} - xe^{-a(x+k)}}{2(\cosh ak - 1)}$
$\phi_{12}(x,a,k)$	$\dfrac{(x+k)e^{-ax} + xe^{-a(x+k)}}{2(\cosh ak + 1)}$

arise in the same way and can be plotted just as easily when the parameters a, b, and k are known

Table 7.2 lists the inverses of all elementary periodic-type transforms which are likely to be encountered. Of course, as Example 5 illustrated, it is usually necessary to employ the method of partial fractions before the results of Table 7 2 can be applied

table 7.2

	Laplace transform		Inverse over general interval $nk < t < (n + 1)k$ $-k < \tau < 0$
1	$\dfrac{1}{s(1 - e^{-ks})}$		$\phi_1(t,k)$
2	$\dfrac{1}{s(1 + e^{-ks})}$		$\phi_2(t,k)$
3	$\dfrac{1}{s^2(1 - e^{-ks})}$		$\phi_3(t,k)$
4	$\dfrac{1}{s^2(1 + e^{-ks})}$		$\phi_4(t,k)$
5	$\dfrac{1}{(s + a)(1 - e^{-ks})}$	$a \neq 0$	$\phi_5(\tau,a,k) - \phi_5(t,a,k)$
6	$\dfrac{1}{(s + a)(1 + e^{-ks})}$	$a \neq 0$	$(-1)^n\phi_6(\tau,a,k) + \phi_6(t,a,k)$
7	$\dfrac{s + a}{[(s + a)^2 + b^2](1 - e^{-ks})}$		$\phi_7(\tau,a,b,k) - \phi_7(t,a,b,k)$†
8	$\dfrac{s + a}{[(s + a)^2 + b^2](1 + e^{-ks})}$		$(-1)^n\phi_8(\tau,a,b,k) + \phi_8(t,a,b,k)$‡
9	$\dfrac{b}{[(s + a)^2 + b^2](1 - e^{-ks})}$		$\phi_9(\tau,a,b,k) - \phi_9(t,a,b,k)$†
10	$\dfrac{b}{[(s + a)^2 + b^2](1 + e^{-ks})}$		$(-1)^n\phi_{10}(\tau,a,b,k) + \phi_{10}(t,a,b,k)$‡
11	$\dfrac{1}{(s + a)^2(1 - e^{-ks})}$	$a \neq 0$	$\phi_{11}(\tau,a,k) - \phi_{11}(t,a,k)$
12	$\dfrac{1}{(s + a)^2(1 + e^{-ks})}$	$a \neq 0$	$(-1)^n\phi_{12}(\tau,a,k) + \phi_{12}(t,a,k)$

† The possibility that, simultaneously, a is zero and bk is an even multiple of π is to be ruled out

‡ The possibility that, simultaneously, a is zero and bk is an odd multiple of π is to be ruled out.

Formulas 1 to 4 are obtained by obvious applications of Theorem 1 and of Theorem 3, Sec. 7.2. Formula 5 was derived in detail in Example 4, and the derivations of Formulas 6 to 10 follow almost exactly the same pattern. All that is necessary is to express as complex exponentials the sines and cosines which appear in the inverses of the individual terms. The expression for $f(t)$ over any interval $nk < t < (n + 1)k$ is then, as in Example 4, just a finite geometric progression which can be summed and converted to a purely real form without difficulty.

The derivation of Formulas 11 and 12 are somewhat different because of the repeated factors in the denominators of the transforms. Over the general interval $nk < t < (n + 1)k$, these lead to expressions for $f(t)$ which are series of the form

$$\sum_{j=0}^{n} (t - jk)e^{-a(t-jk)} = te^{-at} \sum_{j=0}^{n} (e^{ak})^j - ke^{-at} \sum_{j=0}^{n} j(e^{ak})^j$$

in the case of Formula 11, and

$$\sum_{j=0}^{n} (-1)^j(t - jk)e^{-a(t-jk)} = te^{-at} \sum_{j=0}^{n} (-e^{ak})^j - ke^{-at} \sum_{j=0}^{n} j(-e^{ak})^j$$

in the case of Formula 12. In each instance, the second series is not a geometric progression and must be summed by other means. Fortunately, the results of Example 3, Sec. 4.5, are applicable, and through their use the inverses given in Table 7.2 can easily be established.

The transient, or t-evaluated, terms in the inverses in Table 7.2 are all continuous for all $t \geqq 0$. This is true of the periodic, or τ-evaluated, terms if and only if the degree of the polynomial part of the denominator of the transform exceeds the degree of the numerator by more than 1. If this is not the case, there is a jump of 1 at each of the points $t = k, 2k, 3k, \ldots, nk, \ldots$ if the denominator of the transform contains $1 - e^{-ks}$ and a jump of $(-1)^n$ if the denominator of the transform contains $1 + e^{-ks}$.

EXAMPLE 6

A simple series circuit contains the elements $R = 400$, $L = 0.2$, $C = 10^{-6}$. At $t = 0$, while the circuit is completely passive, an exponential "saw-tooth" voltage wave, equal to $E_0e^{-5{,}000t}$ throughout one period and repeating itself every 0.002 sec, is switched into the circuit. Find the total current and also the steady-state current which result.

The differential equation to be solved is

$$0.2\frac{di}{dt} + 400i + 10^6 \int_0^t i\,dt = E(t)$$

Taking the Laplace transform of both sides, we obtain

$$\mathcal{L}\{i\}\left(0.2s + 400 + \frac{10^6}{s}\right) = E_0 \frac{\int_0^{0.002} e^{-5{,}000t}e^{-st}\,dt}{1 - e^{-0.002s}} = \frac{E_0}{1 - e^{-0.002s}}\left[\frac{e^{-t(s+5{,}000)}}{-(s + 5{,}000)}\right]_0^{0.002}$$

or

$$\mathcal{L}\{i\}\,\frac{s^2+2{,}000s+5\times 10^6}{5s} = E_0\,\frac{1-e^{-0.002s-10}}{(s+5{,}000)(1-e^{-0.002s})}$$

$$= E_0\,\frac{(1-e^{-10})+e^{-10}(1-e^{-0.002s})}{(s+5{,}000)(1-e^{-0.002s})}$$

$$= E_0\,\frac{e^{-10}}{s+5{,}000} + E_0\,\frac{1-e^{-10}}{(s+5{,}000)(1-e^{-0.002s})}$$

Hence,

$$\mathcal{L}\{i\} = \frac{5E_0e^{-10}s}{(s+5{,}000)[(s+1{,}000)^2+(2{,}000)^2]} + \frac{5E_0(1-e^{-10})s}{(s+5{,}000)[(s+1{,}000)^2+(2{,}000)^2](1-e^{-0.002s})}$$

Now by simple partial-fraction manipulations we find

$$\frac{s}{(s+5{,}000)[(s+1{,}000)^2+(2{,}000)^2]} = \frac{1}{4{,}000}\left[-\frac{1}{s+5{,}000}+\frac{s+1{,}000}{(s+1{,}000)^2+(2{,}000)^2}\right]$$

From this point the entire solution can be written down at once

$$i = \frac{5E_0e^{-10}}{4{,}000}(-e^{-5{,}000t}+e^{-1{,}000t}\cos 2{,}000t)$$
$$-\frac{5E_0(1-e^{-10})}{4{,}000}[\phi_5(\tau, 5{,}000, 0.002)-\phi_5(t, 5{,}000, 0.002)]$$
$$+\frac{5E_0(1-e^{-10})}{4{,}000}[\phi_7(\tau, 1{,}000, 2{,}000, 0.002)-\phi_7(t, 1{,}000, 2{,}000, 0.002)]$$

The steady-state current is described by the terms in τ

$$i_{ss} = -\frac{5E_0(1-e^{-10})}{4{,}000}[\phi_5(\tau, 5{,}000, 0.002)-\phi_7(\tau, 1{,}000, 2{,}000, 0.002)]$$

or written out at length:

$$i_{ss} = -\frac{E_0(1-e^{-10})}{800}\left[\frac{e^{-5{,}000\tau}}{e^{10}-1}-\frac{e^{-1{,}000\tau}\cos 2{,}000(\tau+0.002)-e^{-1{,}000(\tau+0.002)}\cos 2{,}000\tau}{2(\cosh 2-\cos 4)}\right]$$

This function, plotted for $-0.002<\tau<0$, defines one complete cycle of the steady-state current. Of course, the unit jumps in ϕ_5 and ϕ_7 at the ends of each period just cancel, leaving the steady-state current continuous, as, of course, it must be.

The operational solution of a problem such as this, leading as it does to a relatively simple, finite expression for the response, is in general to be preferred to the use of Fourier series, which leaves the answer in the form of an infinite series.

EXERCISES

1. Using Theorem 1, verify that

a. $\mathcal{L}\{\sin bt\} = b/(s^2+b^2)$ b. $\mathcal{L}\{\cos bt\} = s/(s^2+b^2)$

2. Obtain the Laplace transform of the staircase function (Fig. 7.14a) by direct evaluation of the Laplace transform integral.

Find the Laplace transforms of the periodic functions whose definitions over one period are:

3. $f(t) = \sin t \qquad 0<t<\pi$

4. $f(t) = \begin{cases} \sin t & 0<t<\pi \\ 0 & \pi<t<2\pi \end{cases}$

5. $f(t) = \begin{cases} t & 0<t<a \\ 0 & a<t<2a \end{cases}$

6. $f(t) = \begin{cases} 1 & 0<t<a \\ 0 & a<t<2a \\ -1 & 2a<t<3a \\ 0 & 3a<t<4a \end{cases}$

Find the inverse of each of the following transforms:

7 $\dfrac{s}{(s+1)(s+2)(s^2+1)(1-e^{-2s})}$

8 $\dfrac{e^{-s}}{s(s^2+2s+5)(1+e^{-s})}$

9 $\dfrac{1}{(s-1)(s+2)^2(1+e^{-s})}$

10 $\dfrac{3s+5}{(s+1)(s^2+4s+5)(1-e^{-3s})}$

Solve the following differential equations and explain your answers, $f(t)$ being in each case a periodic function defined over one period as indicated

11 $y' + 4y + 3\int_0^t y\,dt = f(t)$ $\quad f(t) = \begin{cases} 1 & 0 < t < 2 \\ -1 & 2 < t < 4 \end{cases}$ $\quad y_0 = 1$

12 $y'' + 4y' + 4y = f(t)$ $\quad f(t) = \begin{cases} 1 & 0 < t < 1 \\ 0 & 1 < t < 2 \end{cases}$ $\quad y_0 = y_0' = 0$

13 $y'' + y = f(t)$ $\quad f(t) = \begin{cases} 1 & 0 < t < \pi \\ 0 & \pi < t < 2\pi \end{cases}$ $\quad y_0 = y_0' = 0$

14 According to the footnotes to Table 7 2, certain values of a, b, and k cannot be allowed to occur simultaneously in Formulas 7, 8, 9, and 10 of Table 7 2 because the formulas become meaningless for these values Why is this?

15 Derive each of the following formulas of Table 7 2

a Formula 6 b Formula 7 c Formula 8

d Formula 9 e Formula 10 f Formula 11

g Formula 12

7.7

Convolution and the Duhamel formulas

We shall conclude this chapter by establishing a result concerning the product of transforms which is of considerable theoretical as well as practical interest

THEOREM 1

$$\mathcal{L}\{f(t)\}\mathcal{L}\{g(t)\} = \mathcal{L}\left\{\int_0^t f(t-\lambda)g(\lambda)\,d\lambda\right\}$$
$$= \mathcal{L}\left\{\int_0^t f(\lambda)g(t-\lambda)\,d\lambda\right\}$$

PROOF Working with the term on the right in the first equality, we have by definition

(1) $$\mathcal{L}\left\{\int_0^t f(t-\lambda)g(\lambda)\,d\lambda\right\} = \int_0^\infty \left[\int_0^t f(t-\lambda)g(\lambda)\,d\lambda\right] e^{-st}\,dt$$

Now $$u(t-\lambda) = \begin{cases} 1 & \lambda < t \\ 0 & \lambda > t \end{cases}$$

and thus $$f(t-\lambda)g(\lambda)u(t-\lambda) = \begin{cases} f(t-\lambda)g(\lambda) & \lambda < t \\ 0 & \lambda > t \end{cases}$$

Since this product vanishes for all values of λ greater than t, the inner integration in (1) can be extended to infinity if the factor $u(t-\lambda)$ is inserted in the integrand. Hence,

$$(2)\qquad \mathcal{L}\left\{\int_0^t f(t-\lambda)g(\lambda)\,d\lambda\right\} = \int_0^\infty \left[\int_0^\infty f(t-\lambda)g(\lambda)u(t-\lambda)\,d\lambda\right] e^{-st}\,dt$$

Now our usual assumptions about the functions we transform are sufficient to permit the order of integration in (2) to be interchanged:

$$(3)\qquad \begin{aligned}\mathcal{L}\left\{\int_0^t f(t-\lambda)g(\lambda)\,d\lambda\right\} &= \int_0^\infty \left[\int_0^\infty f(t-\lambda)g(\lambda)u(t-\lambda)e^{-st}\,dt\right] d\lambda \\ &= \int_0^\infty g(\lambda)\left[\int_0^\infty f(t-\lambda)u(t-\lambda)e^{-st}\,dt\right] d\lambda\end{aligned}$$

Because of the presence of $u(t-\lambda)$, the integrand of the inner integral is identically zero for all $t < \lambda$. Hence, the inner integration effectively starts not at $t = 0$, but at $t = \lambda$. Therefore

$$(4)\qquad \mathcal{L}\left\{\int_0^t f(t-\lambda)g(\lambda)\,d\lambda\right\} = \int_0^\infty g(\lambda)\left[\int_\lambda^\infty f(t-\lambda)e^{-st}\,dt\right] d\lambda$$

Now, in the inner integral on the right of (4), let $t-\lambda = \tau$ and $dt = d\tau$. Then

$$\begin{aligned}\mathcal{L}\left\{\int_0^t f(t-\lambda)g(\lambda)\,d\lambda\right\} &= \int_0^\infty g(\lambda)\left[\int_0^\infty f(\tau)e^{-s(\tau+\lambda)}\,d\tau\right] d\lambda \\ &= \int_0^\infty g(\lambda)e^{-s\lambda}\left[\int_0^\infty f(\tau)e^{-s\tau}\,d\tau\right] d\lambda \\ &= \left[\int_0^\infty f(\tau)e^{-s\tau}\,d\tau\right]\left[\int_0^\infty g(\lambda)e^{-s\lambda}\,d\lambda\right] \\ &= \mathcal{L}\{f(t)\}\mathcal{L}\{g(t)\} \qquad \text{as asserted}\end{aligned}$$

From symmetry, the second form of the theorem can be obtained by interchanging $f(t)$ and $g(t)$.

The **convolution, or Faltung,*** **integral**

$$\int_0^t f(t-\lambda)g(\lambda)\,d\lambda$$

is frequently denoted simply by $f(t)*g(t)$. In this symbolism Theorem 1 becomes

$$\mathcal{L}\{f\}\mathcal{L}\{g\} = \mathcal{L}\{f*g\} = \mathcal{L}\{g*f\}$$

EXAMPLE 1

If $\mathcal{L}\{f(t)\} = 1/(s^2+4s+13)^2$, what is $f(t)$?

Clearly, we can write $\mathcal{L}\{f(t)\}$ in the form

$$1/[(s+2)^2+3^2]^2$$

and then use the corollary of the first shifting theorem (Theorem 5, Sec. 7.4) to obtain

$$(5)\qquad f(t) = \mathcal{L}^{-1}\left\{\frac{1}{[(s+2)^2+3^2]^2}\right\} = e^{-2t}\mathcal{L}^{-1}\left\{\frac{1}{(s^2+3^2)^2}\right\}$$

Now

$$\frac{1}{(s^2+3^2)^2} = \mathcal{L}\left\{\frac{\sin 3t}{3}\right\}\mathcal{L}\left\{\frac{\sin 3t}{3}\right\}$$

---

* German for *folding*

Hence, by the convolution theorem,

$$\begin{aligned}\mathcal{L}^{-1}\left\{\frac{1}{(s^2+3^2)^2}\right\} &= \frac{1}{9}\int_0^t \sin 3(t-\lambda)\sin 3\lambda\,d\lambda \\ &= \frac{1}{9}\int_0^t \frac{\cos(6\lambda-3t)-\cos 3t}{2}\,d\lambda \\ &= \frac{1}{18}\left[\frac{\sin(6\lambda-3t)}{6}-\lambda\cos 3t\right]_0^t \\ &= \frac{1}{18}\left(\frac{\sin 3t}{3}-t\cos 3t\right)\end{aligned}$$

Therefore, from (5),

$$f(t) = \frac{e^{-2t}(\sin 3t - 3t\cos 3t)}{54}$$

This example illustrates how in certain cases the convolution theorem can be used in place of a fourth Heaviside theorem to handle repeated quadratic factors in the denominator of a transform

EXAMPLE 2

Find a particular integral of the differential equation

$$y'' + 2ay' + (a^2+b^2)y = f(t)$$

Taking the Laplace transform of the given equation, assuming $y_0 = y_0' = 0$, since we desire only a *particular* solution, we find

$$\mathcal{L}\{y\} = \frac{1}{(s+a)^2+b^2}\mathcal{L}\{f(t)\}$$

Now

$$\frac{1}{(s+a)^2+b^2} = \mathcal{L}\left\{\frac{e^{-at}\sin bt}{b}\right\}$$

Hence

$$\mathcal{L}\{y\} = \mathcal{L}\{f(t)\}\,\mathcal{L}\left\{\frac{e^{-at}\sin bt}{b}\right\}$$

and thus, by the convolution theorem,

$$y = \frac{1}{b}\int_0^t f(t-\lambda)e^{-a\lambda}\sin b\lambda\,d\lambda$$

or, equally well,

$$y = \frac{1}{b}\int_0^t f(\lambda)e^{-a(t-\lambda)}\sin b(t-\lambda)\,d\lambda = \frac{e^{-at}}{b}\int_0^t f(\lambda)e^{a\lambda}\sin b(t-\lambda)\,d\lambda$$

It is interesting to compare this procedure with the method of variation of parameters (Sec 2 4) for the determination of particular integrals of linear differential equations The two give identical results in the case of constant-coefficient linear differential equations

An especially important application of the convolution theorem makes it possible to determine the response of a system to a general excitation if its response to a unit step function is known To develop this idea we shall need the concepts of *transfer function* and *indicial admittance.*

Any physical system capable of responding to an excitation can be thought of as a device by means of which an input function is transformed into an output function. If we assume that all initial conditions are zero at the moment when a single excitation, or **input,** $f(t)$ begins to act, then, by setting up the differential equations describing the system, taking Laplace transforms, and solving for the transform of the **output** $y(t)$, we obtain a relation of the form

$$\mathcal{L}\{y(t)\} = \frac{\mathcal{L}\{f(t)\}}{Z(s)} \tag{6}$$

where $Z(s)$ is a function of s whose coefficients depend solely on the parameters of the system Moreover, in the usual applications to linear systems, $Z(s)$ will be just the quotient of two polynomials in s

In electrical problems where the input is an applied voltage $E_0e^{j\omega t}$ and the output is the resultant current, the function $Z(s)$, except for the fact that the frequency variable $j\omega$ is replaced by the Laplace transform parameter s, is just the impedance of the network However, the importance of $Z(s)$ is not restricted to electrical circuits, and for systems of all sorts the function

$$\frac{1}{Z(s)} = \frac{\mathcal{L}\{y(t)\}}{\mathcal{L}\{f(t)\}} = \frac{\mathcal{L}\{\text{output}\}}{\mathcal{L}\{\text{input}\}}$$

is an exceedingly important quantity, usually called the **transfer function**. In particular, after s has been replaced by $j\omega$, the transfer function can be used to determine the effect of any system on the phase and amplitude of a sinusoidal input of arbitrary frequency, just as in the electrical case

If a unit step function is applied to a system with transfer function $1/Z(s)$, then from (6) we have

$$\mathcal{L}\{y(t)\} = \frac{\mathcal{L}\{u(t)\}}{Z(s)} = \frac{1}{sZ(s)}$$

The response in this particular case is called the **indicial admittance** $A(t)$; that is,

$$\mathcal{L}\{A(t)\} = \frac{1}{sZ(s)} \tag{7}$$

Using (7) we can now rewrite (6) in the form

$$\mathcal{L}\{y(t)\} = \frac{\mathcal{L}\{f(t)\}}{Z(s)} = \frac{s\mathcal{L}\{f(t)\}}{sZ(s)} = s\mathcal{L}\{A(t)\}\mathcal{L}\{f(t)\}$$

Hence, by the convolution theorem,

$$\mathcal{L}\{y(t)\} = s\mathcal{L}\left\{\int_0^t A(t-\lambda)f(\lambda)\,d\lambda\right\} = s\mathcal{L}\left\{\int_0^t A(\lambda)f(t-\lambda)\,d\lambda\right\}$$

But from Theorem 3, Sec 7.4, it follows that

$$y(t) = \frac{d}{dt}\left[\int_0^t A(t-\lambda)f(\lambda)\,d\lambda\right] = \frac{d}{dt}\left[\int_0^t A(\lambda)f(t-\lambda)\,d\lambda\right]$$

Therefore, performing the indicated differentiations,* we have equivalently

(8) $$y(t) = \int_0^t A'(t-\lambda)f(\lambda)\,d\lambda + A(0)f(t)$$

and

(9) $$y(t) = \int_0^t A(\lambda)f'(t-\lambda)\,d\lambda + A(t)f(0)$$

Since $A(t)$ is by definition the response of a system which is initially passive, it follows that $A(0) = 0$. Hence, Eq. (8) becomes simply

(10) $$y(t) = \int_0^t A'(t-\lambda)f(\lambda)\,d\lambda$$

Finally, by making the change of variable $\tau = t - \lambda$ in the integrals in (9) and (10), we obtain the related expressions

(11) $$y(t) = \int_0^t A'(\tau)f(t-\tau)\,d\tau$$

(12) $$y(t) = A(t)f(0) + \int_0^t A(t-\tau)f'(\tau)\,d\tau$$

Formulas (9) to (12) all serve to express the response of a system to a general driving function $f(t)$ in terms of the experimentally accessible response to a unit step function. They are often referred to collectively as **Duhamel's formulas,** after the French mathematician J. M. C. Duhamel (1797–1872).

It is possible to interpret these integrals in physical terms as follows. Let the driving function $f(t)$ be given, and imagine it approximated by a series of step functions, as shown in Fig. 7.18. The first step function is of noninfinitesimal magnitude

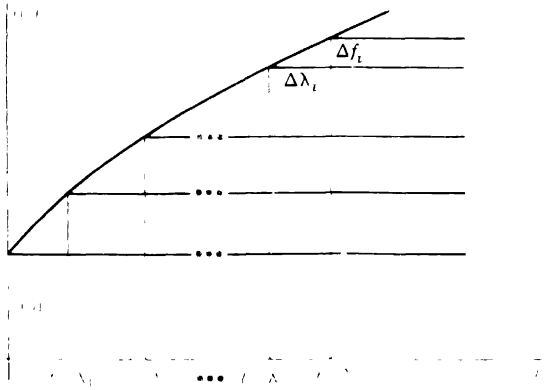


FIGURE 7.18
Plot showing the synthesis of a general function by means of step functions

---

* According to **Leibnitz's rule,** if $F(t) = \int_{a(t)}^{b(t)} \phi(x,t)\,dx$, where a and b are differentiable functions of t and where $\phi(x,t)$ and $\frac{\partial\phi(x,t)}{\partial t}$ are continuous in x and t, then

$$\frac{dF}{dt} = \int_{a(t)}^{b(t)} \frac{\partial\phi(x,t)}{\partial t}\,dx + \phi[b(t),t]\frac{db(t)}{dt} - \phi[a(t),t]\frac{da(t)}{dt}$$

$f(0)$ All later step functions in the approximation are of infinitesimal magnitude, and their contributions in the limit will have to be taken into account by integration Specifically, since

$$\frac{\Delta f}{\Delta \lambda} \doteq \left.\frac{df}{dt}\right|_{t=\lambda} = f'(\lambda)$$

we have for the height Δf_i of the general infinitesimal step function the approximate expression

$$\Delta f_i \doteq f'(\lambda_i)\, \Delta\lambda_i$$

Now if $A(t)$ is the indicial admittance of the system, the first step function $f(0)u(t)$ produces a response equal to

$$f(0)A(t)$$

from the very definition of the indicial admittance as the response per unit excitation For the second step function $\Delta f_1\, u(t - \lambda_1)$, there is a lag of $t = \lambda_1$ units of time before it begins to act Hence the infinitesimal response it produces is

$$\Delta f_1 A(t - \lambda_1) \qquad \text{or} \qquad f'(\lambda_1)\, \Delta\lambda_1\, A(t - \lambda_1)$$

Similarly, the third step function produces the response

$$f'(\lambda_2)\, \Delta\lambda_2\, A(t - \lambda_2)$$

and in general the $(i + 1)$st step function produces the response

$$f'(\lambda_i)\, \Delta\lambda_i\, A(t - \lambda_i)$$

If these contributions to the total response are added, we obtain for the response at a general time t

$$\begin{aligned} y(t) &= f(0)A(t) + f'(\lambda_1)\, \Delta\lambda_1\, A(t - \lambda_1) + f'(\lambda_2)\, \Delta\lambda_2\, A(t - \lambda_2) + \\ &\qquad + f'(\lambda_i)\, \Delta\lambda_i\, A(t - \lambda_i) + \\ &= f(0)A(t) + \Sigma f'(\lambda_i)A(t - \lambda_i)\, \Delta\lambda_i \end{aligned}$$

the summation extending over all the step functions which have begun to act up to the instant t In the limit when $\Delta\lambda_i$ approaches zero and the height of each step function after the first, $f(0)u(t)$, approaches zero, the sum in the last expression becomes an integral, and, except for the dummy variable, we have Eq. (12)

To give a physical interpretation of Eq (10), we must first determine the significance of the derivative of the indicial admittance, $A'(t)$ To do this, we shall need the concept of a *unit impulse*.

Suppose that we have the function shown in Fig. 7 19 This consists of a suddenly applied excitation of constant magnitude acting for a certain period of time and then suddenly ceasing, the product of duration and magnitude being unity If a is very small, the period of application is correspondingly small but the magnitude of the excitation is very great It is sometimes convenient to pursue this idea to the limit and imagine a forcing function of arbitrarily large magnitude acting for an infinitesimal time, the product of duration and intensity

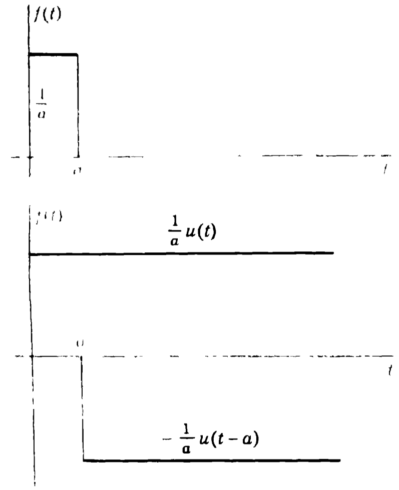


FIGURE 7.19
Plot suggesting the nature of a unit impulse

remaining unity as $a \rightarrow 0$ The resulting "function" is usually referred to as the **unit impulse** $I(t)$ or the **δ function** $\delta(t)$ †

In somewhat different terms, the δ function $\delta(t - t_0)$ is often described by the following purported definition·

$$\delta(t - t_0) = \begin{cases} 0 & t \neq t_0 \\ \infty & t = t_0 \end{cases} \tag{13a}$$

$$\int_{-\infty}^{\infty} \delta(t - t_0)\, dt = 1 \tag{13b}$$

Taken literally this is nonsense, for the area under a curve which coincides with the t-axis at every point but one must surely be zero and not unity, as (13b) asserts However if (13) is considered to be merely suggestive of the limiting process by which we first described the unit impulse, then, whatever its shortcomings as a definition, it is at least as meaningful as certain other useful and reasonably "respectable" concepts in applied mathematics

Consider, for instance, the familiar concept of a concentrated load on a beam (Fig 7 20a) Clearly, such a load is physically unrealizable and must be viewed as an idealization of the following nature Imagine that over the interval $(x_0, x_0 + a)$ the beam bears a distributed load whose magnitude per unit length is P/a (Fig 7 20b) Then no matter how small a may be, the total load on the beam, being equal to the product of the intensity P/a and the interval length a, is just P As $a \rightarrow 0$, the ideal concept of a concentrated load thus emerges as the limiting form of a realizable distributed load If one were now asked to describe

---

† More specifically, $\delta(t)$ is often called the **Dirac δ function, after the British** theoretical physicist P A M Dirac (1902-)

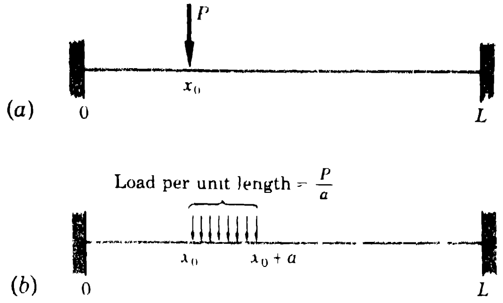


FIGURE 7.20 Plot suggesting the interpretation of a concentrated load on a beam as a unit impulse

the load per unit length $w(x)$ in the limiting case, one would probably give the following "definition".

$$w(x) = \begin{cases} 0 & x \neq x_0 \\ \infty & x = x_0 \end{cases} \tag{14a}$$

$$\int_0^L w(x)\,dx = P \tag{14b}$$

which corresponds in all essential respects to the description of the δ function provided by (13)

One interesting and important property of the δ function is its ability to isolate or reproduce a particular value of a function $f(t)$ according to the following formula

$$\int_{-\infty}^{\infty} f(t)\delta(t - t_0)\,dt = f(t_0) \tag{15}$$

To justify this we revert to the prelimiting approximation to the δ function and use it in place of $\delta(t - t_0)$ in (15) This gives us the approximating integral

$$\int_{t_0}^{t_0+a} f(t)\,\frac{1}{a}\,dt$$

Now, by the *law of the mean for integrals*,* this integral is equal to

$$a\left[\frac{f(\xi)}{a}\right] = f(\xi) \qquad t_0 < \xi < t_0 + a \tag{16}$$

Now as $a \to 0$, perforce $\xi \to t_0$, and so, from (16), the integral approaches $f(t_0)$, as asserted

The unit impulse is only the first of an infinite sequence of so-called **singularity functions.** As a direct generalization of the unit impulse we have the **unit doublet** (Fig 7 21), defined (loosely) as

$$\lim_{a\to 0} \frac{u(t) - 2u(t - a) + u(t - 2a)}{a^2}$$

---

* This asserts that, if $f(t)$ is continuous over the closed range of integration $a \leq t \leq b$, then there exists at least one value of t, say $t = \xi$, between a and b such that

$$\int_a^b f(t)\,dt = (b - a)f(\xi)$$

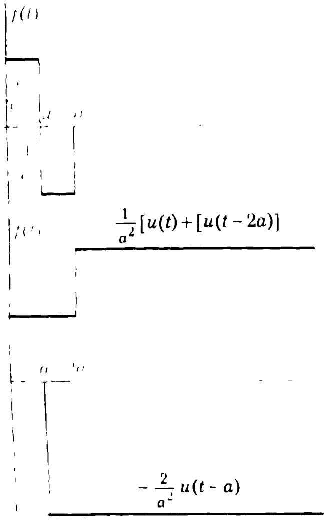


FIGURE 7.21 Plot suggesting the nature of a unit doublet

the **unit triplet**, defined similarly as

$$\lim_{a\to 0}\frac{u(t)-3u(t-a)+3u(t-2a)-u(t-3a)}{a^3}$$

and so on, indefinitely Some of the properties of these "functions" will be found among the exercises at the end of this section

It is interesting and important that in many applications the use of the δ function can be rigorously justified by arguments based on what is known as the *Stieltjes integral*,* a generalization of the familiar Riemann integral More generally, the singularity functions are examples of mathematical objects known as *generalized functions*, or *distributions*, which are studied in the recently developed *theory of distributions* †

To determine the Laplace transform of a unit impulse, we return to the prelimiting approximation

$$\frac{u(t)-u(t-a)}{a}$$

shown in Fig 7 19 Transforming this expression, we have, for all $a > 0$,

$$\frac{1}{a}\left(\frac{1}{s}-\frac{e^{-as}}{s}\right)=\frac{1-e^{-as}}{as}$$

As $a \to 0$, this transform assumes the indeterminate form 0/0, but, evaluating it in the usual way by L'Hospital's rule, we obtain immediately the limiting value 1 In the same way we can show that the transforms of the unit doublet and the unit triplet are, respectively, s and s^2, and the transforms of the other singularity functions follow exactly the same pattern Since these transforms do not approach zero as s becomes infinite, we know from Corollary 1 of Theorem 5, Sec. 7 1, that they are not the transforms of piecewise regular functions of exponential order This, of course, is obvious, for although the singularity functions are all of exponential order, they are limiting forms involving unbounded behavior in the neighborhood of the origin and hence are not piecewise regular

---

* Named for the Dutch mathematician T J Stieltjes (1856–1894)

† An introductory account of the theory of distributions can be found in Athanasios Papoulis, "The Fourier Integral and Its Applications," pp 269–282, McGraw-Hill Book Company, New York, 1962.

We are now in a position to resume our attempt to give a physical interpretation to Formula (10) For convenience let us denote by $h(t)$ the response of the system under discussion when the driving function is a unit impulse We have already seen [Eq (6)] that

$$\mathcal{L}\{y(t)\} = \frac{\mathcal{L}\{f(t)\}}{Z(s)}$$

Hence, if $f(t)$ is a unit impulse, so that $\mathcal{L}\{f(t)\} = 1$ and $y(t) = h(t)$, we have

$$\mathcal{L}\{h(t)\} = \frac{1}{Z(s)} = s\,\frac{1}{sZ(s)} = s\mathcal{L}\{A(t)\}$$

Thus, from Theorem 3, Sec 7 4, it follows that

$$h(t) = \frac{dA(t)}{dt} = A'(t)$$

or, in words, *the response of a system to a unit impulse is the derivative of the response of the system to a unit step function*

Now let $f(t)$, in the general case, be approximated by a series of infinitesimal impulses, as shown in Fig 7 22 For the first

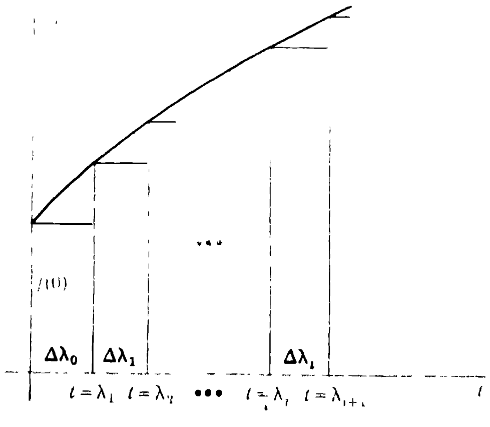


FIGURE 7.22
Plot showing the synthesis of a general function by means of impulses

impulse, whose magnitude by definition is the product

$$f(0)\,\Delta\lambda_0 \equiv f(\lambda_0)\,\Delta\lambda_0$$

the infinitesimal response is $[f(\lambda_0)\,\Delta\lambda_0]A'(t)$, since $A'(t) \equiv h(t)$ is the response per unit impulse. The second impulse does not occur until $t = \lambda_1$, hence the response it produces is $[f(\lambda_1)\,\Delta\lambda_1]A'(t - \lambda_1)$, and in general, the response produced by the $(i + 1)$st impulse is

$$[f(\lambda_i)\,\Delta\lambda_i]A'(t - \lambda_i)$$

If these contributions to the total response are added, we obtain for the response at a general time t

$$y(t) = \Sigma f(\lambda_i)A'(t - \lambda_i)\,\Delta\lambda_i$$

the summation extending over all impulses which have acted on the system up to the time t. In the limit when each $\Delta\lambda \to 0$, the last sum becomes an integral, and we have Formula (10)

EXERCISES

Find the inverse of each of the following transforms

1 $\dfrac{1}{(s^2 + 4)^2}$

2 $\dfrac{s}{(s^2 + 9)^3}$

3 $\dfrac{s}{s + 2}$

4 $\dfrac{s^4 + 2s + 3}{s^2 + 4}$

5 $\dfrac{s^2 + 4s + 4}{(s^2 + 4s + 13)^2}$

6 $\dfrac{s^4 + 2s^2 - s}{(s + 1)(s^2 + 1)^2}$

7 Using the convolution formula, find a particular integral of the equation

$$y'' + 2ay' + a^2y = f(t)$$

8 Using the convolution formula, find a particular integral of the equation

$$y'' + (a + b)y' + aby = f(t)$$

9 Verify that the Laplace transform of the unit doublet is s and that the Laplace transform of the unit triplet is s^2

10 If $D(t)$ denotes the unit doublet function, show that

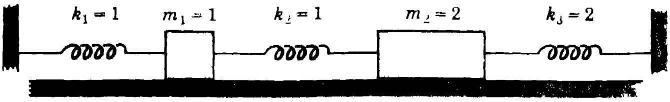
$$\int_{-\infty}^{\infty} f(t)D(t - t_0)\,dt = -f'(t_0)$$

11 Find $A(t)$ and $h(t)$ for the equation $y'' + 3y' + 2y = 0$, verify that $h(t) = A'(t)$, and then verify Formulas (10) and (12) when this equation is "driven" by the function $f(t) = e^t$

12 **a** Find $A(t)$ and $h(t)$ for the system shown in Fig 7 23 if the input is applied to m_1 and if the output is the response, i e , displacement, of m_2 Verify that $h(t) = A'(t)$ What is the response of m_2 to an arbitrary force $f(t)$ applied to m_1 when the system is at rest in its equilibrium position?

b Work part **a** if the input is applied to m_2 and the output is the response of m_1

FIGURE 7.23



13 Find $A(t)$ and $h(t)$ for the system shown in Fig 7.24 if the input is applied across the indicated terminals and if the output is the current through R_2 Verify that $h(t) = A'(t)$ What is the current through R_2 due to an arbitrary voltage $E(t)$ applied across the terminals when all charges and currents in the system are zero?

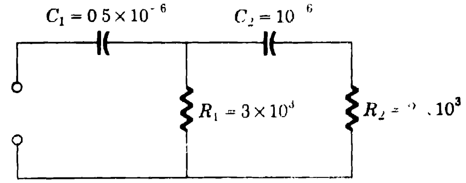

$C_1 = 0\ 5 \times 10^{-6}$
$C_2 = 10^{-6}$
$R_1 = 3 \times 10^3$
$R_2 =$ [illegible] $\times 10^3$

FIGURE 7 24

14 Show that the solution of the equation $ay'' + by' + cy = 0$ ($y_0 = 0$, $y'_0 = 1$) is exactly the same as the solution of the equation $ay'' + by' + cy = a\delta(t)$ ($y_0 = y'_0 = 0$) Does this fact have a physical interpretation? With what combination of singularity functions must an initially passive, second-order equation be driven in order to have the same solution as the undriven equation with initial conditions $y = y_0$, $y' = y'_0$?

15 Show that $f(t)*[g(t)*h(t)] = \int_0^t \int_0^\lambda f(t - \lambda)g(\lambda - \mu)h(\mu)\, d\mu\, d\lambda$

16 Show that $f(t)*[g(t)*h(t)] = [f(t)*g(t)]*h(t)$ and that

$$f(t)*[g(t) \pm h(t)] = [f(t)*g(t)] \pm [f(t)*h(t)]$$

17 Show that $\mathcal{L}\{f(t)\}\mathcal{L}\{g(t)\}\mathcal{L}\{h(t)\} = \mathcal{L}\{f(t)*g(t)*h(t)\}$

18 Show that $1*1 = t$ and that $1*1*1 = t^2/2$ What is the generalization of these results to n factors?

19 Evaluate (a) $\delta(t - a)*f(t)$, (b) $u(t - a)*f(t)$, (c) t^m*t^n if m and n are nonnegative integers

20 If $f(0) = g(0) = 0$, show that $f'(t)*g(t) = f(t)*g'(t)$ and that

$$[f(t)*g(t)]' = \frac{f'(t)*g(t) + f(t)*g'(t)}{2}$$

CHAPTER EIGHT

Partial Differential Equations

8.1

Introduction

In our previous work we have seen how the analysis of mechanical and electrical systems containing lumped parameters often leads to ordinary differential equations However, assumptions to the effect that all masses exist as mass points, that all springs are weightless, or that the elements of an electrical circuit are concentrated in ideal resistances, inductances, and capacitances rather than continuously distributed are frequently not sufficiently accurate In such cases a more realistic approach usually leads to one or more partial differential equations which must be solved to obtain a description of the behavior of the system In this chapter we shall discuss such equations as they commonly arise in engineering and in physics We shall begin our study by examining in detail the derivation from physical principles of certain typical partial differential equations Then, knowing the forms of most frequent occurrence, we shall investigate methods of solution and their application to specific problems

8.2

The derivation of equations

One of the first problems to be attacked through the use of partial differential equations was that of the vibration of a stretched, flexible string Today, after nearly 250 years, it is still an excellent initial example

Let us consider, then, an elastic string, stretched under a tension T between two points on the x-axis (Fig. 8 1a) The weight of the string per unit length after it is stretched we suppose to be a known function $w(x)$. Besides the elastic and inertia forces inherent in the system, the string may also be acted upon by a distributed load whose magnitude per unit length we assume to be a known function of x, y, t, and the transverse velocity y, say $f(x,y,y,t)$. In formulating the problem we assume that

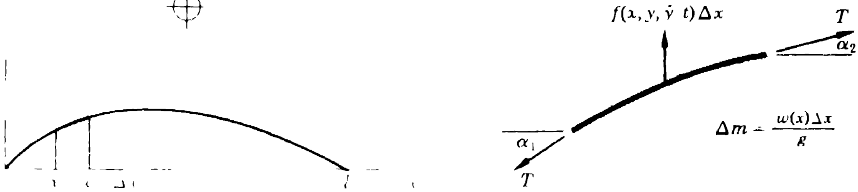
a The motion takes place entirely in one plane, and in this plane each particle moves at right angles to the equilibrium position of the string

b The deflection of the string during the motion is so small that the resulting change in length of the string has no effect on the tension T.

c The string is perfectly flexible, i e, can transmit force only in the direction of its length

d The slope of the deflection curve of the string is at all points and at all times so small that with satisfactory accuracy $\sin \alpha$ can be replaced by $\tan \alpha$, where α is the inclination angle of the tangent to the deflection curve

Gravitational and frictional forces, if any, we suppose to be taken into account in the expression for the load per unit length $f(x,y,y,t)$


$f(x, y, \dot{y}\ t)\Delta x$
T
α_2
$\Delta m = \frac{w(x)\Delta x}{g}$
α_1
T

(*a*) (*b*)

FIGURE 8.1
A typical element of a vibrating string

With these assumptions in mind, let us consider a general infinitesimal segment of the string as a free body (Fig 8 1*b*) By assumption **a**, the mass of such an element is $\Delta m = w(x)\,\Delta x/g$ By assumption **b**, the forces which act at the ends of the element are the same, namely, T By assumption **c**, these forces are directed along the respective tangents to the deflection curve; and, by assumption **d**, their transverse components are

$$T \sin \alpha_2 = T \sin \alpha \bigg|_{x+\Delta x} \doteq T \tan \alpha \bigg|_{x+\Delta x}$$

and $$T \sin \alpha_1 = T \sin \alpha \bigg|_{x} \doteq T \tan \alpha \bigg|_{x}$$

The acceleration produced in Δm by these forces and by the portion of the distributed load $f(x,y,y,t)\ \Delta x$ which acts over the interval Δx is approximately $\frac{\partial^2 y}{\partial t^2}$, where y is the ordinate of an arbitrary point of the element. The time derivative is here **written as a partial derivative because obviously y depends not** only upon t but upon x as well. Applying Newton's law to the element, we can thus write

(1)
$$\frac{w(x)\,\Delta x}{g}\frac{\partial^2 y}{\partial t^2} = T\tan\alpha\Big|_{x+\Delta x} - T\tan\alpha\Big|_{x} + f(x,y,y,t)\,\Delta x$$

or, dividing by Δx,

$$\frac{w(x)}{g}\frac{\partial^2 y}{\partial t^2} = T\left(\frac{\tan\alpha\Big|_{x+\Delta x} - \tan\alpha\Big|_{x}}{\Delta x}\right) + f(x,y,y,t)$$

The fraction on the right-hand side consists of the difference between $\tan\alpha$ at $x + \Delta x$ and at x, divided by the difference Δx In other words, it is precisely the difference quotient for the function $\tan\alpha$ Hence its limit as $\Delta x \to 0$ is the derivative of $\tan\alpha$ with respect to x, that is, $\dfrac{\partial \tan\alpha}{\partial x}$. But, since $\tan\alpha = \dfrac{\partial y}{\partial x}$, this can be written simply as $\dfrac{\partial^2 y}{\partial x^2}$. Our final result, then, is that the deflection $y(x,t)$ of a stretched string satisfies the partial differential equation*

(2)
$$\frac{\partial^2 y}{\partial t^2} = \frac{Tg}{w(x)}\frac{\partial^2 y}{\partial x^2} + \frac{g}{w(x)}f(x,y,y,t)$$

In most important applications the weight of the string per unit length $w(x)$ is a constant, and there are no external forces, i e , $f(x,y,\dot{y},t)$ is identically zero When this is the case, Eq (2) reduces to the **one-dimensional wave equation**

(3)
$$\frac{\partial^2 y}{\partial t^2} = a^2\frac{\partial^2 y}{\partial x^2} \qquad a^2 = \frac{Tg}{w}$$

The dimensions of a^2 are

$$\frac{\text{Force} \times \text{acceleration}}{\text{Weight/unit length}} = \frac{(ML/T^2)(L/T^2)}{(ML/T^2)(1/L)} = \frac{L^2}{T^2}$$

that is, a has the dimensions of velocity The significance of this will become apparent in Sec 8.3 when we discuss the D'Alembert solution of the wave equation

Closely related to the vibrating string is the vibrating membrane To obtain the partial differential equation describing its behavior, we suppose that it is stretched across some closed curve C in the x,y-plane and that when it vibrates each particle moves in a direction perpendicular to the x,y-plane We assume, further, that the weight per unit area of the membrane after it is

---

* The question of what constitutes a satisfactory derivation of the partial differential equation describing a given physical system is not a simple one To attempt to give a careful limiting argument is, in effect, "to strain at a gnat and swallow a camel," since, being ultimately atomic, no physical system is continuous Perhaps our purported derivations should be regarded merely as plausibility arguments suggesting that certain partial differential equations be accepted as the axioms of a theoretical or "rational" study of applied mathematics, whose practical importance, in contrast to its purely mathematical interest, is to be judged by how well its conclusions describe past observations and predict new ones

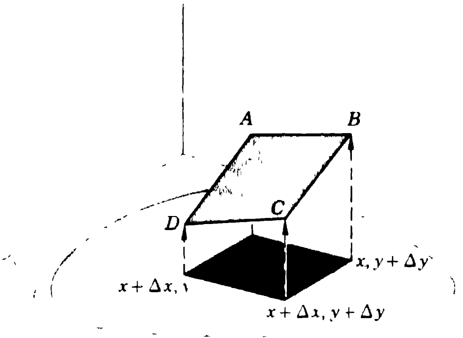


FIGURE 8.2
A typical element of a vibrating membrane

stretched is a known function $w(x,y)$ and that the tension per unit length is the same at all points and in all directions Finally, we suppose that the membrane is acted upon by a known distributed force whose magnitude per unit area is $f(x,y,z,\dot{z},t)$ Then by computing the transverse, or z-components, of the tensile forces acting across the boundaries of a typical two-dimensional element of the membrane (Fig 8 2) and applying Newton's law to the mass of such an element, we find without difficulty that the deflection of the membrane $z(x,y,t)$ satisfies the equation

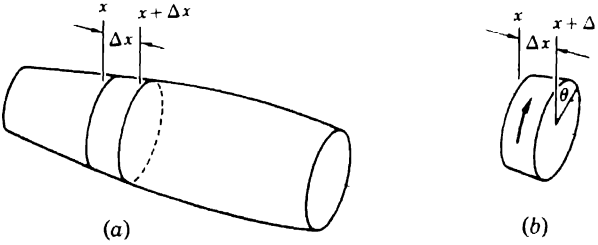
$$\frac{\partial^2 z}{\partial t^2} = \frac{Tg}{w(x,y)}\left(\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}\right) + \frac{g}{w(x,y)} f(x,y,z,\dot{z},t) \tag{4}$$

If the membrane is uniform and if there are no external forces, i e , if $f(x,y,z,\dot{z},t) \equiv 0$, then Eq (4) reduces to the **two-dimensional wave equation**

$$\frac{\partial^2 z}{\partial t^2} = a^2\left(\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}\right) \qquad a^2 = \frac{Tg}{w} \tag{5}$$

Here, as in the case of the vibrating string, the parameter a has the dimensions of velocity

As a third problem leading to a partial differential equation, let us consider a shaft vibrating torsionally (Fig 8.3a) The material of the shaft we assume to have a modulus of elasticity in shear E_s and to be of uniform weight per unit volume ρ The cross-section area of the shaft at a distance x from one end we suppose to be a known function, say $A(x)$ The polar moment of inertia $J(x)$ of a general cross section about its center of gravity



(a) (b)

FIGURE 8.3
A typical element of a vibrating shaft

we also suppose known. In addition to the obvious elastic and inertia torques, the shaft may also be acted upon by a distributed torque whose magnitude per unit length is a known function, say $f(x,\theta,\dot{\theta},t)$, where θ is the angle through which a general cross section has rotated from its equilibrium position and $\dot{\theta}$ is the angular velocity with which that cross section rotates while the shaft is vibrating We assume further that

a All cross sections of the shaft remain plane during rotation
b Each cross section rotates about its center of gravity
c The shape of a general cross section does not depart greatly from a circle.

Frictional torques, if any, we suppose to be taken into account in the expression for the distributed torque per unit length, $f(x,\theta,\dot{\theta},t)$.

We begin by considering as a free body an infinitesimal segment of the shaft bounded by two cross sections a distance Δx apart (Fig. 8.3*b*). The mass of such a disk is approximately

$$\Delta m = \frac{\rho A(x)\,\Delta x}{g}$$

and its radius of gyration is

$$k = \sqrt{\frac{J(x)}{A(x)}}$$

Hence, its polar moment of inertia is approximately

$$\Delta I = k^2\,\Delta m = \frac{J(x)}{A(x)}\cdot\frac{\rho A(x)\,\Delta x}{g} = \frac{J(x)\rho\,\Delta x}{g}$$

The rotation of such an element is produced by the portion of the distributed torque $f(x,\theta,\dot{\theta},t)\,\Delta x$ which acts on it and by the torque T, transmitted to it through the end sections by the adjacent portions of the shaft Therefore, applying Newton's law in torsional form, we have

$$\frac{J(x)\rho\,\Delta x}{g}\frac{\partial^2\theta}{\partial t^2} = T\Big|_{x+\Delta x} - T\Big|_{x} + f(x,\theta,\dot{\theta},t)\,\Delta x$$

or, dividing by Δx and then letting $\Delta x \to 0$,

$$\frac{J(x)\rho}{g}\frac{\partial^2\theta}{\partial t^2} = \frac{\partial T}{\partial x} + f(x,\theta,\dot{\theta},t) \tag{6}$$

Now, from strength of materials, we recall that the torque transmitted through any cross section of a twisted shaft is proportional to the twist per unit length, i.e , the slope of the (θ,x)-curve at that cross section.

$$T = k\frac{\partial\theta}{\partial x}$$

The proportionality constant k is known as the **torsional rigidity**. For shafts which are solids of revolution it can be shown that

$$k = E_sJ(x)$$

and this result can be used with satisfactory accuracy whenever the cross sections of a shaft are approximately circular. Hence, in such cases Eq. (6) becomes

(7) $$\frac{J(x)\rho}{g}\frac{\partial^2\theta}{\partial t^2} = \frac{\partial\left[E_s J(x)\dfrac{\partial\theta}{\partial x}\right]}{\partial x} + f(x,\theta,\dot{\theta},t)$$

However, for configurations whose cross sections differ appreciably from circles, it is necessary to determine the torsional rigidity k by experimental means and continue the solution of Eq. (6) by numerical rather than analytical methods.

In most elementary applications the shafts are of uniform circular cross section and there are no external, distributed torques. In such cases $J(x)$ is a constant, $f(x,\theta,\dot{\theta},t)$ is identically zero, and Eq. (7) therefore reduces to

(8) $$\frac{\partial^2\theta}{\partial t^2} = a^2\frac{\partial^2\theta}{\partial x^2} \qquad a^2 = \frac{E_s g}{\rho}$$

which is again just the one-dimensional wave equation.

Another vibration problem of considerable practical interest concerns the transverse vibrations of a beam. To obtain the partial differential equation describing these vibrations, let us first choose a coordinate system such that the beam in its undeflected position coincides with a portion of the x-axis and the deflections occur in the direction of the y-axis A general cross section of the beam we assume to be of known area $A(x)$ and known moment of inertia $I(x)$ about its neutral axis. The material of the beam we suppose to be of weight per unit volume ρ and modulus of elasticity E. In addition to the intrinsic elastic and inertia forces, the beam may also be acted upon by a distributed load of known intensity $f(x,y,\dot{y},t)$. Gravitational and frictional forces, if any, we suppose included in this distributed load. Finally, we assume that all particles of the beam move in a purely transverse direction, i.e., that the slight rotation of the cross sections as the beam vibrates is negligible

Now from the discussion in Sec. 2.6 we recall the following formulas of beam flexure:

$$M(x) = EI(x)\frac{d^2y}{dx^2} \qquad \frac{dM(x)}{dx} = V(x) \qquad \frac{dV(x)}{dx} = -w(x)$$

where $M(x)$ = bending moment at a general cross section
$V(x)$ = shear, or net transverse force, to the right of a general cross section
$w(x)$ = load per unit length at a general cross section

Hence, combining these relations into a single equation, we have

(9) $$w(x) = -\frac{\partial V(x)}{\partial x} = -\frac{\partial^2 M(x)}{\partial x^2} = -\frac{\partial^2\left[EI(x)\dfrac{\partial^2 y}{\partial x^2}\right]}{\partial x^2}$$

where the derivatives are now written as partial derivatives, since in our problem y depends upon t as well as upon x

During vibration the load per unit length on the beam consists of two parts: the external load $f(x,y,\dot{y},t)$ and the inertia load due to the motion of the beam itself. Now the mass of an infinitesimal segment of the beam of length Δx is approximately $[\rho A(x)\, \Delta x]/g$, and the transverse acceleration of such a mass element is $\frac{\partial^2 y}{\partial t^2}$. Hence the inertia load per unit length is

$$\frac{\frac{\rho A(x)\, \Delta x}{g} \frac{\partial^2 y}{\partial t^2}}{\Delta x} = \frac{\rho A(x)}{g} \frac{\partial^2 y}{\partial t^2}†$$

and, therefore, the total load per unit length is

$$w(x) = \frac{\rho}{g} A(x) \frac{\partial^2 y}{\partial t^2} + f(x,y,y,t)$$

Substituting this into Eq (9), we have finally

(10) $$\frac{\partial^2 \left[EI(x) \frac{\partial^2 y}{\partial x^2} \right]}{\partial x^2} = -\frac{\rho}{g} A(x) \frac{\partial^2 y}{\partial t^2} - f(x,y,y,t)$$

In many important applications the beam under consideration is of constant cross section and there is no external load, that is, A and I are constants and $f(x,y,y,t) \equiv 0$. Under these conditions Eq (10) reduces to the simpler form

(11) $$a^2 \frac{\partial^4 y}{\partial x^4} = -\frac{\partial^2 y}{\partial t^2} \qquad a^2 = \frac{EIg}{A\rho}$$

In this case the parameter a does *not* have the dimensions of velocity

An entirely different class of problems leading to partial differential equations is encountered in the study of the flow of heat in conducting regions To obtain the equation governing this phenomenon we must make use of the following experimental facts.

a Heat flows in the direction of decreasing temperature

b The rate at which heat flows through an area is proportional to the area and to the temperature gradient normal to the area

c The quantity of heat gained or lost by a body when its temperature changes is proportional to the mass of the body and to the temperature change.

---

† The sign of the inertia load per unit length can be checked by observing that, when the beam is instantaneously concave toward the positive y-axis, its elements are either losing velocity in the negative y-direction or gaining velocity in the positive y-direction and, hence, have positive acceleration Therefore the inertia load per unit length is positive, as required by the convention we established in Sec. 2 6 (Fig 2 2) Similarly, when the beam is instantaneously convex toward the positive y-axis, the acceleration of its particles is negative, and so, too, is the inertia load per unit length.

The proportionality constant in **b** is called the **thermal conductivity** of the material k. The proportionality constant in **c** is called the **specific heat** c.

Let us now consider the thermal conditions in an infinitesimal element of a conducting solid (Fig 8.4). If the weight of the conducting material per unit volume is ρ, the mass of such an element is

$$\Delta m = \frac{\rho\,\Delta x\,\Delta y\,\Delta z}{g}$$

Then, if Δu is the temperature change which occurs in the interval Δt, the quantity of heat stored in the element in this time is, by **c**,

$$\Delta H = c\,\Delta m\,\Delta u = \frac{c\rho\,\Delta x\,\Delta y\,\Delta z\,\Delta u}{g}$$

and the rate at which heat is being stored is approximately

$$\frac{\Delta H}{\Delta t} = \frac{c\rho}{g}\,\Delta x\,\Delta y\,\Delta z\,\frac{\Delta u}{\Delta t} \tag{12}$$

The heat which produces the temperature change Δu comes from two sources In the first place, heat may be generated throughout the body, by electrical or chemical means for instance, at a known rate per unit volume, say $f(x,y,z,t)$. The rate at which heat is being received by the element from this source is, then,

$$f(x,y,z,t)\,\Delta x\,\Delta y\,\Delta z \tag{13}$$

In the second place, the element may also gain heat by virtue of heat transfer through its various faces

In particular, the rate at which heat flows into the element through the rear face $EFGH$ is, by b, approximately

$$-k\,\Delta y\,\Delta z\,\frac{\partial u}{\partial x}\bigg|_{\substack{x \\ y+\frac{1}{2}\Delta y \\ z+\frac{1}{2}\Delta z}}$$

where, as an average figure, we have used the temperature gradient $\partial u/\partial x$ at the mid-point of the face $(x,\ y + \frac{1}{2}\,\Delta y,\ z + \frac{1}{2}\,\Delta z)$. The minus sign is necessary because the element *gains* heat through the rear face if the normal temperature gradient,

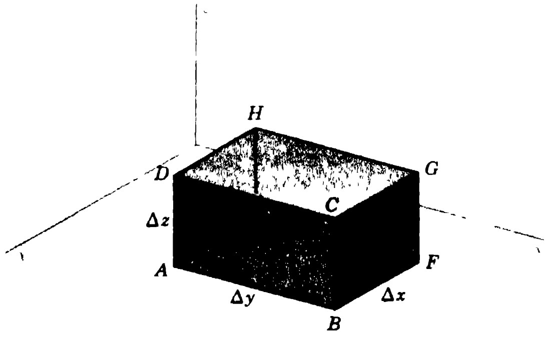


FIGURE 8.4
A typical volume element in a region of three-dimensional heat flow.

i.e., the rate of change of temperature in the x-direction, is *negative*. Similarly the element gains heat through the front face at the approximate rate

$$k\,\Delta y\,\Delta z\,\frac{\partial u}{\partial x}\bigg|_{\substack{x+\Delta x\\ y+\frac{1}{2}\Delta y\\ z+\frac{1}{2}\Delta z}}$$

The sum of these two expressions is the net rate at which the element is gaining heat because of heat flow in the x-direction.

In the same way we find that the rates at which the element gains heat because of flow in the y- and z-directions are, respectively,

$$-k\,\Delta x\,\Delta z\,\frac{\partial u}{\partial y}\bigg|_{\substack{x+\frac{1}{2}\Delta x\\ y\\ z+\frac{1}{2}\Delta z}} + k\,\Delta x\,\Delta z\,\frac{\partial u}{\partial y}\bigg|_{\substack{x+\frac{1}{2}\Delta x\\ y+\Delta y\\ z+\frac{1}{2}\Delta z}}$$

and

$$-k\,\Delta x\,\Delta y\,\frac{\partial u}{\partial z}\bigg|_{\substack{x+\frac{1}{2}\Delta x\\ y+\frac{1}{2}\Delta y\\ z}} + k\,\Delta x\,\Delta y\,\frac{\partial u}{\partial z}\bigg|_{\substack{x+\frac{1}{2}\Delta x\\ y+\frac{1}{2}\Delta y\\ z+\Delta z}}$$

Now the rate at which heat is being stored in the element (12) must equal the rate at which heat is being produced in the element (13) plus the rate at which heat is flowing into the element from the rest of the region. Hence we have the approximate relation

$$\begin{aligned}\frac{cp}{g}\,\Delta x\,\Delta y\,\Delta z\,\frac{\Delta u}{\Delta t} = f(x,y,z,t)\,\Delta x\,\Delta y\,\Delta z\\
+ k\,\Delta y\,\Delta z\left(\frac{\partial u}{\partial x}\bigg|_{\substack{x+\Delta x\\ y+\frac{1}{2}\Delta y\\ z+\frac{1}{2}\Delta z}} - \frac{\partial u}{\partial x}\bigg|_{\substack{x\\ y+\frac{1}{2}\Delta y\\ z+\frac{1}{2}\Delta z}}\right)\\
+ k\,\Delta x\,\Delta z\left(\frac{\partial u}{\partial y}\bigg|_{\substack{x+\frac{1}{2}\Delta x\\ y+\Delta y\\ z+\frac{1}{2}\Delta z}} - \frac{\partial u}{\partial y}\bigg|_{\substack{x+\frac{1}{2}\Delta x\\ y\\ z+\frac{1}{2}\Delta z}}\right)\\
+ k\,\Delta x\,\Delta y\left(\frac{\partial u}{\partial z}\bigg|_{\substack{x+\frac{1}{2}\Delta x\\ y+\frac{1}{2}\Delta y\\ z+\Delta z}} - \frac{\partial u}{\partial z}\bigg|_{\substack{x+\frac{1}{2}\Delta x\\ y+\frac{1}{2}\Delta y\\ z}}\right)\end{aligned}$$

Finally, dividing by $k\,\Delta x\,\Delta y\,\Delta z$ and letting Δx, Δy, Δz, and Δt approach zero, we obtain the equation of heat conduction

$$(14)\qquad a^2\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} + \frac{1}{k}f(x,y,z,t) \qquad a^2 = \frac{cp}{kg}$$

The parameter a in this equation does not have the dimensions of velocity.

In many important cases, heat is neither generated nor lost in the body, and we are interested only in the limiting, steady-state temperature distribution when all change of temperature with time has ceased. Under these conditions both $f(x,y,z,t)$ and $\partial u/\partial t$ are identically zero, and Eq. (14) becomes simply

$$(15)\qquad \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0$$

This exceedingly important equation, which arises in many applications besides steady-state heat flow, is known as **Laplace's equation** and is often written in the abbreviated form

(16) $$\nabla^2 u = 0$$

As a final example of the derivation of partial differential equations from physical principles, we consider the flow of electricity in a long cable or transmission line. We assume the cable to be imperfectly insulated so that there is both capacitance and current leakage to ground (Fig. 8 5). Specifically, let

x = distance from sending end of cable
$e(x,t)$ = potential at any point on cable at any time
$i(x,t)$ = current at any point on cable at any time
R = resistance of cable *per unit length*
L = inductance of cable *per unit length*
G = conductance to ground *per unit length of cable*
C = capacitance to ground *per unit length of cable*

Now the potential at Q is equal to the potential at P minus the drop in potential along the element PQ Hence, referring to the equivalent circuit shown in Fig. 8 5*b*,

$$e(x + \Delta x) = e(x) - (R\,\Delta x)i - (L\,\Delta x)\frac{\partial i}{\partial t}$$

or $$e(x + \Delta x) - e(x) \equiv \Delta e = -(R\,\Delta x)i - (L\,\Delta x)\frac{\partial i}{\partial t}$$

or finally, dividing by Δx and then letting Δx approach zero,

(17) $$\frac{\partial e}{\partial x} = -Ri - L\frac{\partial i}{\partial t}$$

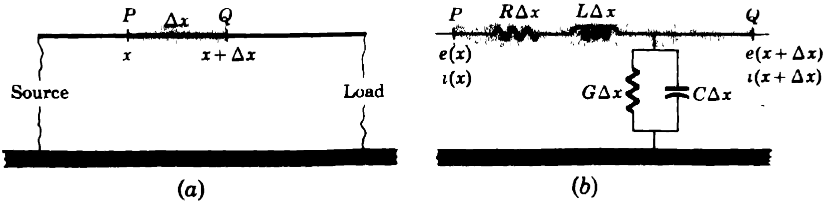
Likewise, the current at Q is equal to the current at P minus the current lost through leakage to ground and the apparent current loss due to the varying charge stored on the element Hence, referring again to Fig 8 5,

$$i(x + \Delta x) = i(x) - (G\,\Delta x)e - (C\,\Delta x)\frac{\partial e}{\partial t}$$

or $$i(x + \Delta x) - i(x) \equiv \Delta i = -(G\,\Delta x)e - (C\,\Delta x)\frac{\partial e}{\partial t}$$

or finally, dividing by Δx and then letting Δx approach zero,

(18) $$\frac{\partial i}{\partial x} = -Ge - C\frac{\partial e}{\partial t}$$



(*a*) (*b*)

FIGURE 8.5
A typical element of a transmission line.

If we differentiate Eq (17) with respect to x and Eq. (18) with respect to t, we obtain

$$\frac{\partial^2 e}{\partial x^2} = -R\frac{\partial i}{\partial x} - L\frac{\partial^2 i}{\partial x\,\partial t}$$

$$\frac{\partial^2 i}{\partial t\,\partial x} = -G\frac{\partial e}{\partial t} - C\frac{\partial^2 e}{\partial t^2}$$

If we eliminate the term $\frac{\partial^2 i}{\partial t\,\partial x}\left(\equiv \frac{\partial^2 i}{\partial x\,\partial t}\right)$ between these two equations and then substitute for $\frac{\partial i}{\partial x}$ from (18), we find that e satisfies the equation

(19) $$\frac{\partial^2 e}{\partial x^2} = LC\frac{\partial^2 e}{\partial t^2} + (RC + GL)\frac{\partial e}{\partial t} + RGe$$

By differentiating Eq (17) with respect to t and Eq (18) with respect to x and then eliminating the derivatives of e, we obtain a similar equation for i.

(20) $$\frac{\partial^2 i}{\partial x^2} = LC\frac{\partial^2 i}{\partial t^2} + (RC + GL)\frac{\partial i}{\partial t} + RGi$$

Equations (19) and (20) are known as the **telephone equations.**

Two special cases of the telephone equations are worthy of note

a If leakage and inductance are negligible, that is, if $G = L = 0$, as they are, for example, for coaxial cables, Eqs (19) and (20) reduce, respectively, to

(21a) $$\frac{\partial^2 e}{\partial x^2} = RC\frac{\partial e}{\partial t}$$

(21b) $$\frac{\partial^2 i}{\partial x^2} = RC\frac{\partial i}{\partial t}$$

These are known as the **telegraph equations.** Mathematically, they are identical with the one-dimensional heat equation, that is, the equation to which (14) reduces when there are no heat sources in the conducting region and the temperature depends only on one space coordinate.

b At high frequencies the factor introduced by the time differentiation is large. Hence the terms involving e and $\frac{\partial e}{\partial t}$ or i and $\frac{\partial i}{\partial t}$ are insignificant in comparison with the terms containing the corresponding second derivatives $\frac{\partial^2 e}{\partial t^2}$ and $\frac{\partial^2 i}{\partial t^2}$. In this case Eqs. (19) and (20) reduce, respectively, to

(22a) $$\frac{\partial^2 e}{\partial x^2} = LC\frac{\partial^2 e}{\partial t^2}$$

(22b) $$\frac{\partial^2 i}{\partial x^2} = LC\frac{\partial^2 i}{\partial t^2}$$

Each of these is an example of the one-dimensional wave equation [Eq. (3)], $1/\sqrt{LC}$ having, in fact, the dimensions of velocity These equations are obtained at any frequency, of course, if $R = G = 0$.

It is interesting to note that nowhere in the derivation of any of the preceding equations was any use made of boundary conditions In other words, the same partial differential equation is satisfied by a vibrating beam, for instance, whether the beam is built-in at one end and free at the other, built-in at both ends, or simply supported at both ends Similarly, the flow of heat in a body is described by the same equation whether the surface is maintained at a constant temperature, insulated against heat loss, or allowed to cool freely by conduction to the surrounding medium In general, as we shall soon see, the role of boundary conditions, for example, permanent conditions of constraint or of temperature, is to determine the *form* of those solutions of a partial differential equation which are relevant to a particular problem Subsequent to this, the initial conditions of displacement, velocity, or temperature, say, determine specific values for the arbitrary constants appearing in these solutions

EXERCISES

1. Supply the details of the derivation of Eq (4) for the transverse vibrations of a membrane
2. What is the form of the heat equation if the thermal conductivity k and the specific heat c vary from point to point in the body?
3. Consider the telephone equations in the so-called distortionless case when $RC = LG$, and put $a^2 = RG$ and $v^2 = 1/LC$ Prove that if $e(x,t)$ [or, equally well, $i(x,t)$] is written in the form $e(x,t) = \epsilon^{-avt}y(x,t)$, then the function y satisfies the wave equation

$$v^2 \frac{\partial^2 y}{\partial x^2} = \frac{\partial^2 y}{\partial t^2}$$

(Note: To avoid confusion with the voltage, ϵ is here used in place of e to denote the base of natural logarithms)

4. Derive the partial differential equation satisfied by the concentration u of a liquid diffusing through a porous solid (Hint. The rate at which liquid diffuses through an area is proportional to the area and to the concentration gradient normal to the area)
5. Consider a region of space filled with a moving fluid Let the density of the fluid at the point (x,y,z) at time t be $\rho(x,y,z,t)$, and let the particle instantaneously at the point (x,y,z) have velocity components v_x, v_y, and v_z, respectively, in the directions of the coordinate axes By considering the flow through an infinitesimal region of dimensions Δx, Δy, Δz, show that the velocity components satisfy the so-called **equation of continuity:**

$$\frac{\partial(\rho v_x)}{\partial x} + \frac{\partial(\rho v_y)}{\partial y} + \frac{\partial(\rho v_z)}{\partial z} + \frac{\partial \rho}{\partial t} = 0$$

6. If $u(x,t)$ is the displacement of a general cross section of a bar which is vibrating longitudinally, show that

$$A(x)\frac{\partial^2 u}{\partial t^2} = \frac{Eg}{\rho}\frac{\partial\left[A(x)\dfrac{\partial u}{\partial x}\right]}{\partial x}$$

where $A(x)$ is the cross-sectional area of the shaft, E is the modulus of elasticity of the material of the shaft, and ρ is the weight per unit volume of the material. (Hint. Use the definition of the modulus of elasticity,

$$E = \frac{\text{stress}}{\text{strain}} = \frac{\text{force/unit area}}{\text{stretch/unit length}}$$

to obtain the expression

$$F = EA\frac{\partial u}{\partial x}$$

for the force transmitted through a general cross section of a stretched bar)

7 a Show that Laplace's equation in three dimensions

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0$$

is satisfied by the function

$$u = \frac{1}{\sqrt{(x-a)^2 + (y-b)^2 + (z-c)^2}}$$

for all values of the constants a, b, c

b Determine whether Laplace's equation in two dimensions

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

is satisfied by the function

$$u = \frac{1}{\sqrt{(x-a)^2 + (y-b)^2}}$$

8 Show that Laplace's equation in two dimensions is satisfied by the function

$$u = \ln\,[(x-a)^2 + (y-b)^2]$$

for all values of the constants a and b

9 Show that, if $z_1(x,y)$ and $z_2(x,y)$ are solutions of the equation

$$p_1(x,y)\frac{\partial^2 z}{\partial x^2} + p_2(x,y)\frac{\partial^2 z}{\partial x\,\partial y} + p_3(x,y)\frac{\partial^2 z}{\partial y^2} + q_1(x,y)\frac{\partial z}{\partial x} + q_2(x,y)\frac{\partial z}{\partial y} + r_1(x,y)z = 0$$

then for all values of the constants c_1 and c_2 the expression $c_1z_1(x,y) + c_2z_2(x,y)$ is also a solution

10 Show that, when Laplace's equation in cartesian coordinates

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0$$

is transformed into cylindrical coordinates by means of the substitutions $x = r\cos\theta$, $y = r\sin\theta$, $z = z$, it becomes

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\partial^2 u}{\partial z^2} = 0$$

8.3

The D'Alembert solution of the wave equation

Each of the partial differential equations we encountered in the last section can be solved by a method of considerable generality known as *separation of variables*. For the one-dimensional wave equation, however, there is also an elegant, special method known as **D'Alembert's solution*** which, because of the importance of this equation, we shall examine in some detail before developing more general techniques

The whole matter is very simple In fact, if f is a function possessing a second derivative, then

$$\frac{\partial f(x - at)}{\partial t} = -af'(x - at) \qquad \frac{\partial f(x - at)}{\partial x} = f'(x - at)$$

$$\frac{\partial^2 f(x - at)}{\partial t^2} = a^2 f''(x - at) \qquad \frac{\partial^2 f(x - at)}{\partial x^2} = f''(x - at)$$

and from these results it is evident that $y = f(x - at)$ satisfies the equation

(1) $$\frac{\partial^2 y}{\partial t^2} = a^2 \frac{\partial^2 y}{\partial x^2}$$

It is an equally simple matter to prove that, if g is an arbitrary twice-differentiable function, then $y = g(x + at)$ is likewise a solution of (1) Hence, since (1) is a linear equation, it follows that the sum

(2) $$y = f(x - at) + g(x + at)$$

is also a solution In fact, it can be shown (see Exercise 10) that if f and g are arbitrary twice-differentiable functions, then (2) is a *complete* solution of (1)

This form of the solution of the wave equation is especially useful for revealing the significance of the parameter a and its dimensions of velocity Suppose, specifically, that we consider the vibrations of a uniform string† stretching from $-\infty$ to ∞ If its transverse displacement is given by (2), we have in fact two waves traveling in opposite directions along the string, each with velocity a. For consider the function $f(x - at)$ At $t = 0$, it defines the curve $y = f(x)$, and at any later time $t = t_1$, it defines the curve $y = f(x - at_1)$ But these two curves are identical except that the latter is translated to the right a distance equal to at_1. Thus the entire configuration moves along the string without distortion a distance of at_1 in t_1 units of time The velocity with which the wave is propagated is therefore

$$v = \frac{at_1}{t_1} = a$$

---

* Named for the French mathematician Jean le Rond D'Alembert (1717–1783) The D'Alembert solution is actually not a special method but rather a special application of a general procedure known as the *method of characteristics* Unfortunately, this cannot be applied with comparable simplicity to problems involving the heat equation and Laplace's equation, and so, despite its theoretical interest, we shall not discuss it here An introduction to the theory can be found in Arnold Sommerfeld, "Partial Differential Equations in Physics," pp. 36–43, Academic Press Inc , New York, 1949

† The use of the string as an illustration is purely a matter of convenience, and *any* quantity satisfying the wave equation possesses the properties developed for the string

Similarly, the function $g(x + at)$ defines a configuration which moves to the left along the string with constant velocity a. The total displacement of the string is, of course, the algebraic sum of these two traveling waves.

To carry the solution through in detail, let us suppose that the initial displacement of the string at any point x is given by $\phi(x)$ and that the initial velocity of the string at any point x is $\theta(x)$. Then, as conditions to determine the form of f and g, we have, from (2) and its first derivative with respect to t,

$$y(x,0) = \phi(x) = f(x) + g(x) \tag{3}$$

$$\left.\frac{\partial y}{\partial t}\right|_{x,0} = \theta(x) = -af'(x) + ag'(x) \tag{4}$$

Dividing Eq. (4) by a and then integrating, we find

$$-f(x) + g(x) = \frac{1}{a}\int_{x_0}^{x} \theta(x)\,dx$$

Combining this with Eq. (3) and introducing the dummy variable s in the integrals, we obtain

$$f(x) = \frac{1}{2}\left[\phi(x) - \frac{1}{a}\int_{x_0}^{x} \theta(s)\,ds\right]$$

$$g(x) = \frac{1}{2}\left[\phi(x) + \frac{1}{a}\int_{x_0}^{x} \theta(s)\,ds\right]$$

With the forms of f and g known, we can now write

$$y = f(x - at) + g(x + at) = \left[\frac{\phi(x - at)}{2} - \frac{1}{2a}\int_{x_0}^{x-at} \theta(s)\,ds\right] + \left[\frac{\phi(x + at)}{2} + \frac{1}{2a}\int_{x_0}^{x+at} \theta(s)\,ds\right]$$

or, combining the integrals,

$$y(x,t) = \frac{\phi(x - at) + \phi(x + at)}{2} + \frac{1}{2a}\int_{x-at}^{x+at} \theta(s)\,ds \tag{5}$$

EXAMPLE 1

A string stretching to infinity in both directions is given the initial displacement

$$\phi(x) = \frac{1}{1 + 8x^2}\,†$$

and released from rest. Determine its subsequent motion.

---

† The initial deflection curve $y = \phi(x)$ clearly violates assumption **d**, Sec. 8.1, since at $x = -\frac{1}{4}$ (for instance), $\phi'(x) \equiv \tan\alpha = \frac{16}{9} = 1.78$ while $\sin\alpha = 0.87$. This difficulty can easily be overcome, however, by assuming instead of $\phi(x)$ a new deflection curve

$$\phi^*(x) = \frac{\phi(x)}{k}$$

where k is a sufficiently large constant, say $k = 10{,}000$. Using $\phi(x)$ instead of $\phi^*(x)$ in this and in similar problems is just a convenient way of eliminating the constant factor $1/k$ at each step of our work.

Since $\theta(x) \equiv 0$, we have from (5) simply

$$y(x,t) = \frac{\phi(x - at) + \phi(x + at)}{2} = \frac{1}{2}\left[\frac{1}{1 + 8(x - at)^2} + \frac{1}{1 + 8(x + at)^2}\right]$$

The deflection of the string when $at = 0.0$, 0.5, and 1.0 is shown in Fig. 8.6

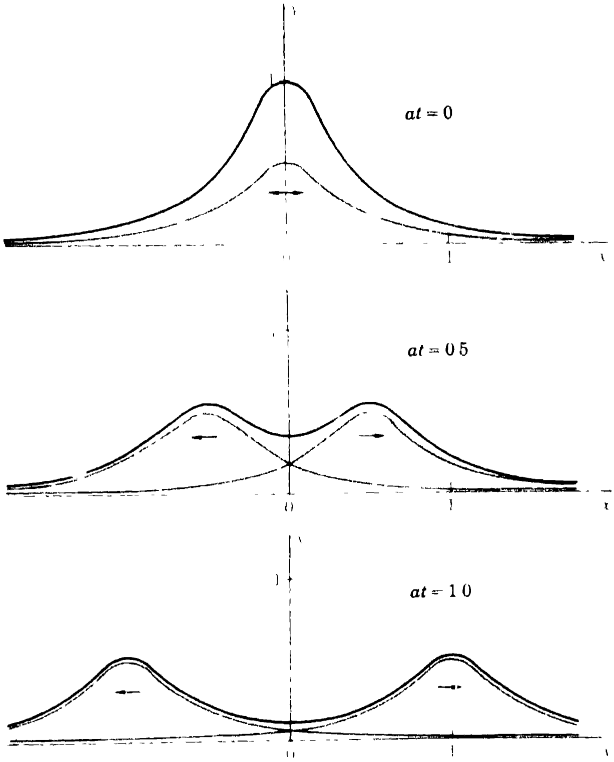


FIGURE 8.6
Plot showing the propagation of a disturbance along a two-way infinite string

The motion of a semi-infinite string whose end is fixed is completely equivalent to the motion of one-half of a two-way infinite string having a fixed point, or **node,** located at some finite point, say the origin To capitalize on this fact we need only imagine the actual string, stretching from 0 to ∞, to be extended in the opposite direction to $-\infty$. The initial conditions of velocity and displacement for the new portion of the string we define to be equal in magnitude but opposite in sign to those given for the actual string * The solution for the resulting two-way infinite string can be written down at once, using Eq (5). In the nature of the extended initial conditions, the displacement at the origin due to the wave traveling to the right from the left half of the string will always be equal but opposite in sign to the displacement at the origin due to the wave traveling to the left from the right half of the string Hence the string will always remain at rest at the origin, and the solution for the right half of the extended string will be precisely the solution of the original problem.

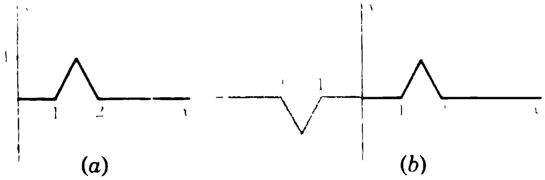
---

* This method of extending the initial conditions is sufficient but not necessary (see Exercise 6).

EXAMPLE 2

A semi-infinite string is given the displacement shown in Fig 8 7a and released from rest Determine its subsequent motion

FIGURE 8.7 A semi-infinite string and its conceptual extension.



We first imagine the string extended to $-\infty$ and released from rest in the extended initial configuration shown in Fig 8 7b Since $\theta(x) \equiv 0$, we have, from (5),

$$y(x,t) = \frac{\phi(x - at) + \phi(x + at)}{2}$$

where $\phi(x)$ is the displacement function shown in Fig 8 7b * We thus have two displacement waves, each of shape defined by $\frac{1}{2}\phi(x)$, one traveling to the right and one traveling to the left along the string Plots of these waves are shown in Fig 8 8 An inspection of these configurations reveals the important fact that a displacement wave is reflected from a fixed or "closed" end without distortion but with reversal of sign

The motion of a finite string can be obtained as the motion of a segment of an infinite string with suitably defined initial displacement and velocity If the string is given between 0 and l, say, we first imagine that it is extended from 0 to $-l$ with initial conditions which are equal but opposite in sign to those for the actual string Then we extend the string to infinity in each direction subject to initial conditions which duplicate with period $2l$ the initial configuration between $-l$ and l

EXAMPLE 3

A string of length l is given the displacement shown in Fig 8 9 and released from rest Determine its subsequent motion

---

* If, as suggested by Fig 8 7a, the graph of $\phi(x)$ has one or more corner points, then, strictly speaking, $\phi(x)$ does not describe an admissible initial displacement function In fact, in the derivation of Eq. (5) both $f(x)$ and $g(x)$ were assumed to be twice differentiable, and, therefore, $\phi(x)$ must also be twice differentiable, which is not the case if there are points where the derivative of $\phi(x)$ is undefined. The apparent solutions obtained from Eq (5) by overlooking this fact are, therefore, at best only formal solutions, and are to be viewed with suspicion unless and until it is verified directly that they satisfy the given partial differential equation and its accompanying boundary and initial conditions Questions concerning the existence and uniqueness of solutions of partial differential equations are quite difficult, and in our work we shall be concerned mainly with techniques for obtaining formal solutions. For an extended discussion of the problem of establishing the validity of solutions derived by purely formal means see, for instance, R V Churchill, "Fourier Series and Boundary Value Problems," 2d ed., pp. 126–163, McGraw-Hill Book Company, New York, 1963.

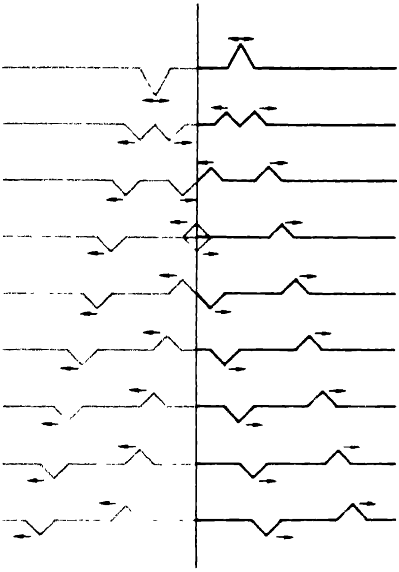
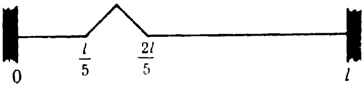


FIGURE 8.8
Plot showing the propagation of a disturbance along a semi-infinite string.

The necessary extension of the string and one half cycle of its motion are shown in Fig 8 10 An inspection of Fig 8 10 shows that the period of the motion, i e , the least time for its return to its initial state, is just the time for either of the traveling waves to traverse a distance $2l$ In other words, since the velocity of the waves is a, the period is $2l/a$ The frequency of the vibrations is therefore $a/2l$ We shall encounter this formula again when we solve the wave equation by the method of separation of variables

FIGURE 8.9
A finite string with initial displacement.



EXERCISES

1 A uniform string stretching from $-\infty$ to ∞ is given the initial displacement

$$y(x,0) = \begin{cases} 1 - |x| & x^2 < 1 \\ 0 & x^2 \geq 1 \end{cases}$$

and released from rest. Find the displacement of the string as a function of x and t, and plot the displacement curves for $at = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$ What is the transverse velocity of the string at $x = 0$?

2 Criticize the following argument "proving" that a string displaced as in Exercise 1 and released from rest will remain motionless. "At $t = 0$ the displacement curve of the string consists exclusively of segments of straight lines, and at all points of any line, or segment of a line, $y = ax + b$, it is obvious that $d^2y/dx^2 = 0$ Hence, at $t = 0$ we have $\frac{\partial^2 y}{\partial x^2} \equiv 0$, and, therefore, from the wave equation

$$\frac{\partial^2 y}{\partial t^2} = a^2 \frac{\partial^2 y}{\partial x^2}$$

it follows that, when $t = 0$, the acceleration $\frac{\partial^2 y}{\partial t^2}$ is zero at all points of the string. But if a particle has zero velocity and if there is no acceleration, i e , if there is no change in velocity, the velocity remains zero Therefore the string will never move "(!)

3 A uniform string stretching from $-\infty$ to ∞ is given the initial displacement

$$y(x,0) = \begin{cases} \cos x & x^2 < \dfrac{\pi^2}{2} \\ 0 & x^2 \geqq \dfrac{\pi^2}{2} \end{cases}$$

and released from rest Find the displacement of the string as a function of x and t, and plot the displacement curves for $at = \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$

4 A uniform string stretching from $-\infty$ to ∞, while at rest in its equilibrium position, is struck in such a way that the portion of the string between $x = -1$ and $x = 1$ is given a velocity of 1 Find the displacement as a function of x and t, and plot the displacement curves for $at = 1$ and 2

5 A uniform string stretching from 0 to ∞ is initially displaced into the curve $y = xe^{-x}$ and released from rest Find its displacement as a function of x and t

6 A uniform string stretching from 0 to ∞ begins its motion with initial displacement $\phi(x)$ and initial velocity $\theta(x)$ Show that its motion can be found as the motion of the right half of a two-way infinite string, provided merely that the initial displacement $\phi(-x)$ and the initial velocity $\theta(-x)$ for the negative extension of the string satisfy the condition

$$\phi(x) + \phi(-x) = -\frac{1}{a}\int_{-x}^{x} \theta(s)\,ds$$

7 If a semi-infinite string begins its motion with initial displacement $\phi(x) = (\sin x)/a$ and initial velocity $\theta(x) = 1$ and if the negative extension of the string is imagined to have the initial displacement $\phi(-x) = 0$, find the necessary initial velocity for the extended portion of the string

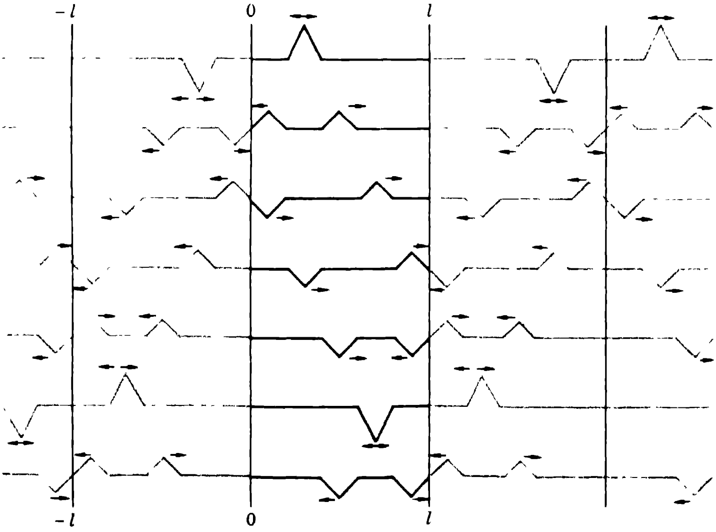


FIGURE 8.10

Plot showing one half cycle of the motion of a finite string

8 The initial displacement of a two-way infinite string is

$$y(x,0) = \frac{1}{1+x^2} \qquad x^2 < \infty$$

With what velocity must the string start to move in order that its subsequent motion will consist solely of a wave traveling to the right?

9 The initial velocity of a two-way infinite string is

$$y(x,0) = \begin{cases} \sin x & x^2 < \pi^2 \\ 0 & x^2 \geq \pi^2 \end{cases}$$

From what initial displacement must the string start to move in order that its subsequent motion will consist solely of a wave traveling to the right?

10 Show that, under the substitutions $u = x - at$ and $v = x + at$, the equation $\dfrac{\partial^2 y}{\partial t^2} = a^2 \dfrac{\partial^2 y}{\partial x^2}$ becomes $\dfrac{\partial^2 y}{\partial u\, \partial v} = 0$ Hence, show that $y = f(x - at) + g(x + at)$ is the most general solution of the one-dimensional wave equation

11 Discuss the possibility of finding solutions of the form $z = f(\lambda x + y)$ for the equation

$$A\frac{\partial^2 z}{\partial x^2} + 2B\frac{\partial^2 z}{\partial x\, \partial y} + C\frac{\partial^2 z}{\partial y^2} = 0 \qquad A,\ B,\ C \text{ constants}$$

and show that, according as $B^2 - AC$ is greater than, equal to, or less than zero, there will be two, one, or no (real) values of λ for which such solutions exist (The given equation is said to be **hyperbolic**, **parabolic**, or **elliptic** in the respective cases, and the nature of its solutions and their properties is significantly different in each case)

12 The equation

$$A(x,y)\frac{\partial^2 z}{\partial x^2} + 2B(x,y)\frac{\partial^2 z}{\partial x\, \partial y} + C(x,y)\frac{\partial^2 z}{\partial y^2} = f\left(z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, x, y\right)$$

is said to be **hyperbolic, parabolic,** or **elliptic** at a point (x,y) according as $B^2(x,y) - A(x,y)C(x,y)$ is, respectively, greater than, equal to, or less than zero For what values of x and y is the equation

$$(1 - y)\frac{\partial^2 z}{\partial x^2} + 2(1 - x)\frac{\partial^2 z}{\partial x\, \partial y} + (1 + y)\frac{\partial^2 z}{\partial y^2} = 0$$

hyperbolic? parabolic? elliptic?

13 Let

$$A(x,y)\frac{\partial^2 z}{\partial x^2} + 2B(x,y)\frac{\partial^2 z}{\partial x\, \partial y} + C(x,y)\frac{\partial^2 z}{\partial y^2} = f\left(z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, x, y\right)$$

and let $\phi(x,y) = c_1$ and $\psi(x,y) = c_2$ be the functions whose derivatives satisfy the equation

$$A(x,y)(w')^2 - 2B(x,y)w' + C(x,y) = 0$$

These functions are called the **characteristics** of the given partial differential equation, and, if $B^2(x,y) - A(x,y)C(x,y) \geq 0$, they define families of (real) curves which are called **characteristic curves.**

a What are the characteristic curves of the one-dimensional wave equation?

b If the given partial differential equation is hyperbolic, show that the change of independent variables defined by the substitutions $u = \phi(x,y)$, $v = \psi(x,y)$ will reduce it to the standard form $\dfrac{\partial^2 z}{\partial u\, \partial v} = F\left(z, \dfrac{\partial z}{\partial u}, \dfrac{\partial z}{\partial v}, u, v\right)$

c If the given partial differential equation is parabolic, show that the change of independent variables defined by the substitutions $u = x$, $v = \phi(x,y)$ will reduce it to the standard form $\dfrac{\partial^2 z}{\partial v^2} = F\left(z, \dfrac{\partial z}{\partial u}, \dfrac{\partial z}{\partial v}, u,v\right)$

d If the given partial differential equation is elliptic, show that the change of independent variables defined by the substitutions $u + iv = \phi(x,y)$, $u - iv = \psi(x,y)$ will reduce it to the standard form $\dfrac{\partial^2 z}{\partial u^2} + \dfrac{\partial^2 z}{\partial v^2} = F\left(z, \dfrac{\partial z}{\partial u}, \dfrac{\partial z}{\partial v}, u,v\right)$

14 Using the substitutions described in the preceding exercise, solve each of the following equations

a $z_{xx} + 3z_{xy} + 2z_{yy} = 0$ **b** $z_{xx} + 4z_{xy} + 4z_{yy} = 0$

c $z_{xx} + 4z_{xy} + 5z_{yy} = 0$ **d** $xz_{xy} + yz_{yy} = 0$

e $xz_{xy} - yz_{yy} = z_y$ **f** $z_{xx} + 2(x + y)z_{xy} + 4xyz_{yy} = 0$

15 **a** Discuss the possibility of extending the D'Alembert solution to the two-dimensional wave equation $a^2(z_{xx} + z_{yy}) = z_{tt}$

b Discuss the possibility of finding solutions of the form $e^{\lambda x + \mu y}$ for the equation

$$Az_{xx} + Bz_{xy} + Cz_{yy} + Dz_x + Ez_y + Fz = 0 \qquad A, B, C, D, E, F \text{ constants}$$

8.4

Separation of variables

We are now ready to consider the solution of partial differential equations by the method of separation of variables Although this method is not universally applicable, it suffices for most of the partial differential equations encountered in elementary applications in engineering and in physics and leads directly to the heart of the branch of mathematics which deals with *boundary value problems*

The idea behind the method is the familiar mathematical stratagem of reducing a new problem to dependence upon an old one In this case we attempt to convert the given partial differential equation into several ordinary differential equations, hopeful that what we know about the latter will prove adequate for a successful continuation

To illustrate the details of the procedure, let us again consider the wave equation, this time taking the torsionally vibrating shaft of finite length as a specific representation

$$\frac{\partial^2 \theta}{\partial t^2} = a^2 \frac{\partial^2 \theta}{\partial x^2}$$

We assume, as a working hypothesis, that solutions for the angle of twist θ exist as products of a function of x alone and a function of t alone

$$\theta(x,t) = X(x)T(t)$$

If this is the case, then partial differentiation of θ amounts to total differentiation of one or the other of the factors of θ, and we have $\frac{\partial^2\theta}{\partial x^2} = X''T$ and $\frac{\partial^2\theta}{\partial t^2} = XT''$.

Substituting these into the wave equation, we obtain

$$XT'' = a^2X''T$$

Dividing by XT then gives

$$\frac{T''}{T} = a^2\frac{X''}{X} \tag{1}$$

as a necessary condition that $\theta(x,t) = X(x)T(t)$ should be a solution

Now the left member of (1) is clearly independent of x Hence (in spite of its appearance) the right-hand side of (1) must also be independent of x, since it is identically equal to the expression on the left Similarly, each member of (1) must be independent of t Therefore, being independent of both x and t, each side of (1) must be a constant, say μ, and we can write

$$\frac{T''}{T} = a^2\frac{X''}{X} = \mu$$

Thus the determination of solutions of the original partial differential equation has been reduced to the determination of solutions of the two ordinary differential equations

$$T'' = \mu T \qquad \text{and} \qquad X'' = \frac{\mu}{a^2}X$$

Assuming that we need consider only real values of μ, there are three cases to investigate

$$\mu > 0 \qquad \mu = 0 \qquad \mu < 0$$

If $\mu > 0$, we can write $\mu = \lambda^2$ In this case the two differential equations and their solutions are

$$T'' = \lambda^2T \qquad X'' = \frac{\lambda^2}{a^2}X$$

$$T = Ae^{\lambda t} + Be^{-\lambda t} \qquad X = Ce^{\lambda x/a} + De^{-\lambda x/a}$$

But a solution of the form

$$\theta(x,t) = X(x)T(t) = (Ce^{\lambda x/a} + De^{-\lambda x/a})(Ae^{\lambda t} + Be^{-\lambda t})$$

cannot describe the undamped vibrations of a system because it is not periodic, i e , does not repeat itself periodically as time increases Hence, although product solutions of the differential equation exist for $\mu > 0$, they have no significance in relation to the problem we are considering

If $\mu = 0$, the equations and their solutions are

$$T'' = 0 \qquad X'' = 0$$

$$T = At + B \qquad X = Cx + D$$

But, again, a solution of the form

$$\theta(x,t) = X(x)T(t) = (Cx + D)(At + B)$$

cannot describe a periodic motion Hence, the alternative $\mu = 0$ must be rejected

Finally, if $\mu < 0$ we can write $\mu = -\lambda^2$. Then the component differential equations and their solutions are

$$T'' = -\lambda^2 T \qquad X'' = -\frac{\lambda^2}{a^2}X$$

$$T = A\cos\lambda t + B\sin\lambda t \qquad X = C\cos\frac{\lambda}{a}x + D\sin\frac{\lambda}{a}x$$

In this case the solution

(2) $$\theta(x,t) = X(x)T(t) = \left(C\cos\frac{\lambda}{a}x + D\sin\frac{\lambda}{a}x\right)(A\cos\lambda t + B\sin\lambda t)$$

is clearly periodic, repeating itself identically every time t increases by $2\pi/\lambda$ In other words, $\theta(x,t)$ represents a vibratory motion with period $2\pi/\lambda$ or frequency $\lambda/2\pi$

It remains now to find the value or values of λ and the constants A, B, C, and D Since the admissible values of λ are determined by the boundary conditions of the problem, the continuation now varies in some respects, depending upon how the shaft is constrained at its ends We shall discuss in turn the following simple cases (Fig 8 11).

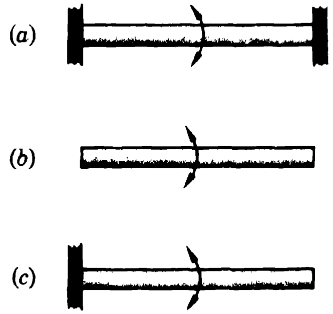


FIGURE 8 11 End conditions for a shaft vibrating torsionally: (a) fixed-fixed, (b) free-free, (c) fixed-free

a Both ends of the shaft are built-in, i.e , are constrained so that no twisting can take place

b Both ends of the shaft are free to twist

c One end of the shaft is built-in; the other is free to twist

If both ends of the shaft are held fixed, we have the following conditions to impose upon the general expression for $\theta(x,t)$, assuming the x-axis chosen along the shaft so that the left end of the shaft is at $x = 0$ and the right end is at $x = l$:

$$\theta(0,t) = \theta(l,t) = 0 \qquad \text{identically in } t$$

Substituting $x = 0$ into the expression (2), we find

$$\theta(0,t) \equiv 0 = C(A\cos\lambda t + B\sin\lambda t)$$

This condition will obviously be fulfilled for all values of t if both A and B are zero In this case, however, $\theta(x,t)$ is zero at all times and the shaft remains motionless, a possible but trivial solution in which we have no interest Hence we are driven to the other alternative, $C = 0$, which reduces (2) to the form

$$\theta(x,t) = D \sin \frac{\lambda}{a} x \,(A \cos \lambda t + B \sin \lambda t)$$

The second boundary condition, namely, that the right end of the shaft remains motionless at all times, requires that

$$\theta(l,t) \equiv 0 = D \sin \frac{\lambda l}{a} (A \cos \lambda t + B \sin \lambda t)$$

As before, we reject the possibility that $A = B = 0$, since it leads only to a trivial solution Moreover, we cannot permit $D = 0$, since that, too, with C already zero, leads to the trivial case The only possibility which remains is that

$$\sin \frac{\lambda l}{a} = 0 \qquad \text{or} \qquad \frac{\lambda l}{a} = n\pi$$

From the continuous infinity of values of the parameter λ for which periodic product solutions of the wave equation exist, we have thus been forced to reject all but the values

$$\lambda_n = \frac{n\pi a}{l} \qquad n = 1, 2, 3, \ldots \tag{3}$$

These and only these values of λ (still infinite in number, however) yield solutions which, in addition to being periodic, also satisfy the end, or boundary, conditions of the problem at hand. With these solutions, one for each admissible value of λ, we must now attempt to construct a solution which will satisfy the remaining conditions of the problem, namely, that the shaft starts its motion at $t = 0$ with a known angle of twist $\theta(x,0) = f(x)$ and a known angular velocity $\left.\frac{\partial \theta}{\partial t}\right|_{x,0} = g(x)$ at every section.

Now the wave equation is linear, and, thus, if we have several solutions, their sum is also a solution Hence, writing the solution associated with the nth value of λ in the form

$$\begin{aligned}\theta_n(x,t) &= \sin \frac{\lambda_n}{a} x \,(A_n \cos \lambda_n t + B_n \sin \lambda_n t) \\ &= \sin \frac{n\pi x}{l} \left(A_n \cos \frac{n\pi a t}{l} + B_n \sin \frac{n\pi a t}{l}\right)^{\dagger}\end{aligned}$$

it is natural enough (though perhaps optimistic, in view of the questions of convergence that are raised) to ask if an *infinite*

---

† The constants A and B now bear subscripts to indicate that they are not necessarily the same in the solutions associated with the different values of λ. The constant D can, of course, be absorbed into the constants A and B and need not be explicitly included.

series of *all* the θ_n's, say

(4)
$$\theta(x,t) = \sum_{n=1}^{\infty} \theta_n(x,t) = \sum_{n=1}^{\infty} \sin\frac{n\pi x}{l}\left(A_n \cos\frac{n\pi at}{l} + B_n \sin\frac{n\pi at}{l}\right)$$

can be made to yield a solution fitting the initial conditions of angular displacement and velocity.

This can be done, and in fact in this case the determination of the coefficients A_n and B_n requires nothing more than a simple application of Fourier series, as developed in Chap 6 For, if we set $t = 0$ in $\theta(x,t)$, we obtain from Eq (4), and the given initial displacement condition,

$$\theta(x,0) \equiv f(x) = \sum_{n=1}^{\infty} A_n \sin\frac{n\pi x}{l}$$

The problem of determining the A_n's so that this will be true is nothing but the problem of expanding a given function $f(x)$ in a half-range sine series over the interval $(0,l)$. Using Theorem 2, Sec 6 4, we have explicitly

$$A_n = \frac{2}{l}\int_0^l f(x) \sin\frac{n\pi x}{l}\,dx$$

Also,
$$\frac{\partial\theta}{\partial t} = \sum_{n=1}^{\infty} \sin\frac{n\pi x}{l}\left(-A_n \sin\frac{n\pi at}{l} + B_n \cos\frac{n\pi at}{l}\right)\frac{n\pi a}{l}$$

Hence, putting $t = 0$, we have from the initial velocity condition,

$$\left.\frac{\partial\theta}{\partial t}\right|_{x,0} \equiv g(x) = \sum_{n=1}^{\infty}\left(\frac{n\pi a}{l} B_n\right)\sin\frac{n\pi x}{l}$$

This, again, merely requires that the B_n's be determined so that the quantities

$$\frac{n\pi a}{l} B_n$$

will be the coefficients in the half-range sine expansion of the known function $g(x)$ Thus

$$\frac{n\pi a}{l} B_n = \frac{2}{l}\int_0^l g(x) \sin\frac{n\pi x}{l}\,dx \qquad \text{or} \qquad B_n = \frac{2}{n\pi a}\int_0^l g(x) \sin\frac{n\pi x}{l}\,dx$$

Aside from convergence questions, our problem is now completely solved We know that a uniform shaft with both ends restrained against twisting can vibrate torsionally at any of an infinite number of natural frequencies,

$$f_n = \frac{\lambda_n}{2\pi} = \frac{na}{2l} \qquad \text{cycles/unit time} \qquad n = 1, 2, 3, \ldots$$

If and when the shaft vibrates at a single one of these frequencies, we know that the angular displacements along the shaft var periodically between extreme values proportional to

$$\sin\frac{n\pi x}{l}$$

Finally, assuming any initial conditions of velocity and displacement which satisfy the Dirichlet conditions, we know how to construct, at least formally,* the instantaneous deflection curve as an infinite series of the deflection curves associated with the respective natural frequencies λ_n

The treatment of the shaft with both ends free follows closely the preceding analysis, once we obtain the proper analytic formulation of the end conditions To obtain this formulation, we observe that at a free end, although we do not know the amount of twist, we do know that there is no torque acting through the end section Recalling from the discussion of Sec 8 2 the expression for the torque transmitted through a general cross section of a twisted shaft, we thus find the free ends characterized by the requirement that

$$E_sJ\left.\frac{\partial\theta}{\partial x}\right|_{\text{end}} = 0$$

Since E_s is a nonzero constant of the material of the shaft and since . cannot vanish for a shaft of uniform section such as we are considering, it follows that at a free end $\frac{\partial\theta}{\partial x} = 0$

Returning to the original product solution (2), we find that

$$\frac{\partial\theta}{\partial x} = \left(-C\frac{\lambda}{a}\sin\frac{\lambda}{a}x + D\frac{\lambda}{a}\cos\frac{\lambda}{a}x\right)(A\cos\lambda t + B\sin\lambda t)$$

Substituting $x = 0$ and equating the result to zero, we obtain the condition

$$\frac{\lambda}{a}D(A\cos\lambda t + B\sin\lambda t) = 0 \qquad \text{for all } t$$

and from this we conclude that $D = 0$ Substituting $x = l$ and again equating to zero, we find

$$-C\frac{\lambda}{a}\sin\frac{\lambda l}{a}(A\cos\lambda t + B\sin\lambda t) = 0$$

Since we cannot permit $C = 0$, we must have

$$\sin\frac{\lambda l}{a} = 0 \qquad \text{or} \qquad \frac{\lambda l}{a} = n\pi$$

Thus, as in the last example, to have the end conditions of the problem fulfilled, λ must be restricted to one of the discrete set of values

$$\lambda_n = \frac{n\pi a}{l} \qquad n = 1, 2, 3, \ldots$$

---

* See the footnote to Example 2, Sec 8 3

Again, we construct the product solution for each admissible value of λ:

$$\begin{aligned}\theta_n(x,t) &= \left(\cos \frac{\lambda_n}{a} x\right)(A_n \cos \lambda_n t + B_n \sin \lambda_n t) \\ &= \cos \frac{n\pi x}{l}\left(A_n \cos \frac{n\pi at}{l} + B_n \sin \frac{n\pi at}{l}\right)\end{aligned}$$

and attempt to form an infinite series of these solutions,

$$\theta(x,t) = \sum_{n=1}^{\infty} \theta_n(x,t) = \sum_{n=1}^{\infty} \cos \frac{n\pi x}{l}\left(A_n \cos \frac{n\pi at}{l} + B_n \sin \frac{n\pi at}{l}\right)$$

which will satisfy the initial displacement condition $\theta(x,0) = f(x)$ and the initial velocity condition $\left.\frac{\partial\theta}{\partial t}\right|_{x,0} = g(x)$

To satisfy the initial displacement condition, we must have

$$\theta(x,0) \equiv f(x) = \sum_{n=1}^{\infty} A_n \cos \frac{n\pi x}{l} \tag{5}$$

which requires that the A_n's be the coefficients in the half-range cosine expansion* of the known function $f(x)$, that is, that

$$A_n = \frac{2}{l}\int_0^l f(x) \cos \frac{n\pi x}{l}\, dx$$

To satisfy the initial velocity condition, we must have

$$\left.\frac{\partial\theta}{\partial t}\right|_{x,0} \equiv g(x) = \sum_{n=1}^{\infty} \left(\frac{n\pi a}{l} B_n\right) \cos \frac{n\pi x}{l} \tag{6}$$

which requires that the quantities

$$\frac{n\pi a}{l} B_n$$

be the coefficients in the half-range cosine series* for $g(x)$ over the interval $(0,l)$, that is, that

$$\frac{n\pi a}{l} B_n = \frac{2}{l}\int_0^l g(x) \cos \frac{n\pi x}{l}\, dx \qquad \text{or} \qquad B_n = \frac{2}{n\pi a}\int_0^l g(x) \cos \frac{n\pi x}{l}\, dx$$

We note in passing that, since the admissible λ's are the same for the free-free shaft and the fixed-fixed shaft, the natural frequencies of the two systems are the same The amplitudes through which they vibrate are not the same, however In fact, for the fixed-fixed shaft we found the distribution of amplitudes along the shaft given by the function $\sin (n\pi x/l)$, whereas for the free-free shaft the amplitudes are given by $\cos (n\pi x/l)$.

The case of the shaft with one end fixed and the other free can be disposed of quickly. Taking the fixed end at $x = 0$ and the free end at $x = l$, we have the two conditions

---

* In general, the half-range cosine expansion of a function begins with a constant term This series does not, because we rejected earlier the possibility $\mu = 0$, which would have led to such a term Had there been an acceptable product solution corresponding to $\mu = 0$ we would, of course, have had to add it to the solutions arising from the assumption $\mu = -\lambda^2$ when we constructed the infinite series for $\theta(x,t)$ (See Exercise 1.)

$$\theta(0,t) = 0 \qquad \text{and} \qquad \left.\frac{\partial\theta}{\partial x}\right|_{l,t} = 0 \qquad \text{for all } t$$

Imposing these upon the general product solution (2) gives

$$C(A \cos \lambda t + B \sin \lambda t) = 0 \qquad \text{or} \qquad C = 0$$

$$\text{and} \qquad \frac{\lambda}{a} D \cos \frac{\lambda l}{a} (A \cos \lambda t + B \sin \lambda t) = 0$$

from which we conclude that

$$\cos \frac{\lambda l}{a} = 0 \qquad \frac{\lambda l}{a} = \frac{(2n-1)\pi}{2} \qquad \text{and finally} \qquad \lambda_n = \frac{(2n-1)a\pi}{2l}$$

The general solution of the problem, formed by adding together the product solutions corresponding to each λ_n, is therefore

$$\theta(x,t) = \sum_{n=1}^{\infty} \sin \frac{\lambda_n}{a} x \, (A_n \cos \lambda_n t + B_n \sin \lambda_n t)$$

$$= \sum_{n=1}^{\infty} \sin \frac{(2n-1)\pi x}{2l} \left[A_n \cos \frac{(2n-1)\pi a t}{2l} + B_n \sin \frac{(2n-1)\pi a t}{2l} \right]$$

To fit the initial displacement condition $\theta(x,0) = f(x)$, we must have

$$f(x) = \sum_{n=1}^{\infty} A_n \sin \frac{(2n-1)\pi x}{2l}$$

This is not quite the usual half-range sine expansion problem, since the arguments of the various terms are not integral multiples of the fundamental argument $\pi x/l$. It is, however, the special half-range sine expansion over $(0,l)$ discussed in Exercise 13, Sec 6.3, where the formula for the coefficients was shown to be

$$A_n = \frac{2}{l} \int_0^l f(x) \sin \frac{(2n-1)\pi x}{2l} \, dx$$

Similarly, to fit the initial velocity condition $\left.\frac{\partial\theta}{\partial t}\right|_{x,0} = g(x)$, we must have

$$g(x) = \sum_{n=1}^{\infty} \left[\frac{(2n-1)\pi a}{2l} B_n \right] \sin \frac{(2n-1)\pi x}{2l}$$

which requires that

$$B_n = \frac{4}{(2n-1)a\pi} \int_0^l g(x) \sin \frac{(2n-1)\pi x}{2l} \, dx$$

EXERCISES

1 Discuss the restrictions implicitly imposed on $f(x)$ and $g(x)$ by the absence of constant terms in the series in Eqs (5) and (6) What is the physical significance of these restrictions?

2 Verify that the solutions of the wave equation obtained in this section can all be written in the form $\theta(x,t) = F(x - at) + G(x + at)$, as required by the D'Alembert theory

3 Which of the following equations can be solved by the method of separation of variables? Where possible, determine the product solutions

a $a\dfrac{\partial^2 u}{\partial x\,\partial y} + bu = 0$ **b** $x^2\dfrac{\partial^2 u}{\partial x^2} + y\dfrac{\partial^2 u}{\partial y^2} = 0$

c $a\dfrac{\partial^2 u}{\partial x^2} + b\dfrac{\partial^2 u}{\partial x\,\partial y} + c\dfrac{\partial u}{\partial y} = 0$ **d** $a\dfrac{\partial^2 u}{\partial x^2} + b\dfrac{\partial^2 u}{\partial y^2} + c\dfrac{\partial^2 u}{\partial z^2} = 0$

e $a\dfrac{\partial^2 u}{\partial x^2} + b\dfrac{\partial^2 u}{\partial x\,\partial y} + c\dfrac{\partial^2 u}{\partial y^2} = 0$ **f** $a\dfrac{\partial^2 u}{\partial x^2} + b\dfrac{\partial^2 u}{\partial y^2} + c\dfrac{\partial u}{\partial x} + d\dfrac{\partial u}{\partial y} = 0$

4 A uniform shaft, fixed at one end and free at the other, is twisted so that each cross section rotates through an angle proportional to the distance from the fixed end If the shaft is released from rest in this position, find its subsequent angular displacement as a function of x and t

5 A uniform shaft, fixed at each end, is twisted so that each cross section rotates through an angle proportional to $x(l - x)$ where l is the length of the shaft and x is the distance from the left end If the shaft is released from rest in this position, find its subsequent angular displacement as a function of x and t

6 A uniform shaft, free at each end, is twisted so that each cross section rotates through an angle proportional to $(2x - l)/2$, where l is the length of the shaft and x is the distance from the left end If the shaft is released from rest in this position, find its subsequent angular displacement as a function of x and t

7 Show that the natural frequencies of a uniform string are given by the formula

$$f_n = \frac{n}{2l}\sqrt{\frac{Tg}{w}} \qquad \text{cycles/unit time}$$

where l is the length of the string, T is the tension under which it is stretched, and w is its weight per unit length How does doubling the tension affect the pitch of the fundamental tone of the string? Why is it that most string instruments either have strings of different lengths or have the lengths of their strings changed by the performer as he plays?

8 A uniform string, stretched between the points (0,0) and (l,0), is given the initial displacement

$$y(x,0) = f(x) = \begin{cases} x & 0 < x < \dfrac{l}{2} \\ l - x & \dfrac{l}{2} < x < l \end{cases}$$

and released from rest Find its subsequent displacement as a function of x and t

9 While in its equilibrium position, a uniform string, stretched between the points (0,0) and (l,0), is given the initial velocity

$$y(x,0) = g(x) = \begin{cases} x & 0 < x < \dfrac{l}{2} \\ l - x & \dfrac{l}{2} < x < l \end{cases}$$

Find its subsequent displacement as a function of x and t

10 While in its equilibrium position, a uniform string, stretched between the points (0,0) and (l,0), is given the initial velocity

$$y(x,0) = g(x) = \begin{cases} 0 & 0 < x < \dfrac{l - k}{2} \\ \dfrac{1}{k} & \dfrac{l - k}{2} < x < \dfrac{l + k}{2} \\ 0 & \dfrac{l + k}{2} < x < l \end{cases}$$

Find its subsequent displacement as a function of x and t Does your answer appear to have a meaningful limit as $k \to 0$? If so, to what problem do you think it is the answer?

11 A uniform string, stretched between the points (0,0) and $(l,0)$ is given the following initial displacement and initial velocity

$$y(x,0) = f(x) = \sin\frac{\pi x}{l} \qquad 0 < x < l$$

$$y(x,0) = g(x) = \begin{cases} 0 & 0 < x < \frac{l}{4} \\ 1 & \frac{l}{4} < x < \frac{3l}{4} \\ 0 & \frac{3l}{4} < x < l \end{cases}$$

Find its subsequent displacement as a function of x and t

12 The curved surface of a rod of length l is perfectly insulated against the flow of heat The rod, which is so thin that heat flow in it can be assumed to be one-dimensional, is initially at the uniform temperature 100° Find the temperature at any point in the rod at any subsequent time if both ends of the rod are kept at the temperature 0° (Hint For heat flow in one dimension, the heat equation reduces to $\dfrac{\partial^2 u}{\partial x^2} = a^2 \dfrac{\partial u}{\partial t}$)

13 Work Exercise 12, with both ends of the rod insulated and the initial temperature distribution in the rod given by

$$u(x,0) = f(x) = u_0\frac{x}{l} \qquad 0 < x < l$$

where x is the distance from the left end of the rod (Hint The temperature gradient through an insulated surface must be 0)

14 Work Exercise 12 with the left end of the rod maintained at the constant temperature 0° and the right end perfectly insulated

15 Show that the torsional vibrations of any uniform fixed-free shaft of length l are always the same as those of the left half of a suitably chosen fixed-fixed shaft of length $2l$ Is the converse true? That is, does the motion of the left half of a fixed-fixed shaft of length $2l$ always represent a possible motion of a fixed-free shaft of length l?

8.5

Orthogonal functions and the general expansion problem

The three examples we considered in the last section embody all the significant features of the general boundary value problem However, they give an exaggerated picture of the role of Fourier series in the final expansion process that is required in order to fit the initial conditions In general, a knowledge of Fourier series, as such, will not suffice to obtain the necessary expansion Hence before we attempt to summarize the major characteristics of boundary value problems, as illustrated in our examples, we shall consider an additional example or two in which Fourier series play no part.

EXAMPLE 1

A slender rod of length l has its curved surface perfectly insulated against the flow of heat Its left end is maintained at the constant temperature $u = 0$, and its right end radiates freely into air of constant temperature $u = 0$. If the initial temperature distribution in the rod is given by

$$u(x,0) = f(x)$$

find the temperature at any point of the rod at any subsequent time.

Since the rod is very thin and since its lateral surface is perfectly insulated, we shall assume that all points of any given cross section are at the same temperature and that the flow of heat in the rod is, therefore, entirely in the x-direction. Thus we have to solve the heat equation [Eq. (14), Sec. 8.2] specialized to one-dimensional flow without heat sources:

$$(1) \qquad \frac{\partial^2 u}{\partial x^2} = a^2 \frac{\partial u}{\partial t}$$

At the left end of the rod we have the obvious fixed-temperature condition $u(0,t) = 0$. At the right end we have a radiation condition which must be formulated analytically before we can proceed with our solution.

Now, according to **Stefan's law**, the amount of heat radiated from a given area dA in a given time interval dt is

$$dQ = \sigma(U^4 - U_0^4)\, dA\, dt$$

where U and U_0 are, respectively, the absolute temperatures of the radiating surface and of the surrounding medium and σ is a proportionality constant. This quantity of heat must have come to the surface by conduction from the interior of the body, hence, we have as a second estimate for dQ the expression

$$dQ = -k \frac{\partial U}{\partial n}\, dA'\, dt$$

where k is the thermal conductivity, $\dfrac{\partial U}{\partial n}$ is the temperature gradient in the direction perpendicular to dA, and dA' is an element of area, congruent to dA, situated in the body an infinitesimal distance from dA in the normal direction. Therefore, equating the two expressions for dQ, we have

$$-k \frac{\partial U}{\partial n}\, dA'\, dt = \sigma(U^4 - U_0^4)\, dA\, dt$$

or, canceling the common factors and expanding $U^4 - U_0^4$ in powers of $U - U_0$,

$$\begin{aligned} -k \frac{\partial U}{\partial n} &= \sigma\{[(U_0 + (U - U_0)]^4 - U_0^4\} \\ &= \sigma[4U_0^3(U - U_0) + 6U_0^2(U - U_0)^2 + \cdots] \end{aligned}$$

Finally, if $U - U_0$ is small in comparison with U_0, as we shall suppose, we can neglect everything on the right except the first term, getting

$$-\frac{\partial U}{\partial n} = h(U - U_0) \qquad h = \frac{4\sigma U_0^3}{k}$$

In our problem, the normal to the surface from which radiation takes place, i.e., the right end of the rod, is the x-axis. Hence if we measure temperatures from U_0 as a reference value, so that $u = U - U_0$, our second boundary condition becomes simply

$$(2) \qquad -\left.\frac{\partial u}{\partial n}\right|_{l,t} \equiv -\left.\frac{\partial u}{\partial x}\right|_{l,t} = hu(l,t)$$

As before, we begin by assuming a product solution $u = XT$ and substituting it into the heat equation (1)

$$X''T = a^2XT'$$

Dividing by XT, we have

$$\frac{X''}{X} = a^2\frac{T'}{T}$$

from which, since x and t are independent variables, we conclude that

$$\frac{X''}{X} \qquad \text{and} \qquad a^2\frac{T'}{T}$$

must equal the same constant, say μ.

If $\mu > 0$, say $\mu = \lambda^2$, we have, from the fraction involving T,

$$T' = \frac{\lambda^2}{a^2}T \qquad \text{and} \qquad T = Ce^{\lambda^2 t/a^2}$$

But this is absurd, since it indicates that the temperature $u = XT$ increases beyond all bounds as t increases. Hence we reject the possibility that $\mu > 0$.

If $\mu = 0$, we have simply

$$X'' = 0 \qquad T' = 0$$

$$X = Ax + B \qquad T = C$$

and, letting $C = 1$, as we can without loss of generality,

$$u = XT = Ax + B$$

For this to be relevant to our problem it must reduce to 0 when $x = 0$, hence $B = 0$. Moreover it must satisfy Eq. (2) when $x = l$, hence $A = 0$. Thus $\mu = 0$ leads only to a trivial solution and must also be rejected.

Finally, if $\mu < 0$, say $\mu = -\lambda^2$, the component differential equations and their solutions are

$$X'' = -\lambda^2 X \qquad T' = -\frac{\lambda^2}{a^2}T$$

$$X = A\cos\lambda x + B\sin\lambda x \qquad T = Ce^{-\lambda^2 t/a^2}$$

and, again letting $C = 1$,

$$u = XT = (A\cos\lambda x + B\sin\lambda x)e^{-\lambda^2 t/a^2}$$

To fit the left end condition we must have $u(0,t) = 0 = Ae^{-\lambda^2 t/a^2}$. Hence $A = 0$, and u reduces to $u = Be^{-\lambda^2 t/a^2}\sin\lambda x$. To fit the right end condition (2), we must have

$$-Be^{-\lambda^2 t/a^2}\lambda\cos\lambda l = hBe^{-\lambda^2 t/a^2}\sin\lambda l$$

or, dividing out the exponential and collecting terms,

$$B(h\sin\lambda l + \lambda\cos\lambda l) = 0$$

If $B = 0$, the solution is trivial. Hence we must have

$$h\sin\lambda l + \lambda\cos\lambda l = 0$$

or $$\tan\lambda l = -\frac{\lambda}{h} = -\frac{\lambda l}{hl}$$

or finally $$\tan z = -\alpha z$$

where $$z = \lambda l \qquad \text{and} \qquad \alpha = \frac{1}{hl}$$

This equation is not like the simple equations

$$\sin\lambda l = 0 \qquad \text{and} \qquad \cos\lambda l = 0$$

which determined the admissible values of λ in the examples of the last section, and its roots cannot be found by inspection. To determine them it is convenient to consider the graphs of the

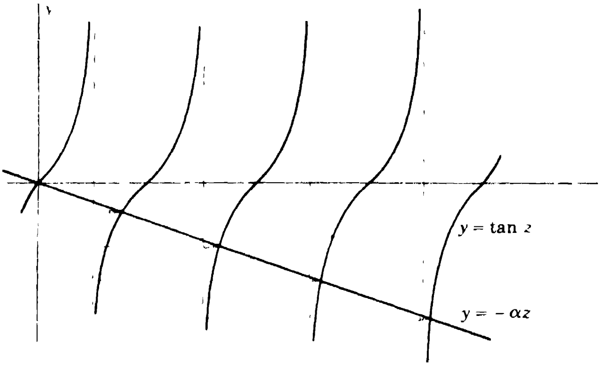


FIGURE 8.12
Plot showing the graphical solution of the equation $\tan z = -\alpha z$

two functions

$$y_1 = \tan z \qquad \text{and} \qquad y_2 = -\alpha z$$

The abscissas of the points of intersection of these curves (Fig. 8.12), being values of z for which $y_1 = y_2$, are then the solutions of the equation

$$\tan z = -\alpha z$$

Obviously, there are an infinite number of roots z_n. However, unlike the roots of $\sin \lambda l = 0$ and $\cos \lambda l = 0$, they are not evenly spaced, although, as the graph in Fig. 8.12 indicates, the interval between successive values of z_n *approaches* π as n becomes infinite.

From each root z_n, we obtain at once the corresponding value of λ

$$\lambda_n = \frac{z_n}{l}$$

and the associated product solution

$$u_n(x,t) = T_n(t)X_n(x) = B_n e^{-\lambda_n^2 t/a^2} \sin \lambda_n x$$

Then we form a series of these particular solutions

$$(3) \qquad u(x,t) = \sum_{n=1}^{\infty} u_n(x,t) = \sum_{n=1}^{\infty} B_n e^{-\lambda_n^2 t/a^2} \sin \lambda_n x$$

and attempt to determine the constants B_n so that the function defined by the series will satisfy the initial condition

$$u(x,0) = f(x)$$

Finally, putting $t = 0$ in (3), we find that this requires

$$(4) \qquad u(x,0) = f(x) = \sum_{n=1}^{\infty} B_n \sin \lambda_n x$$

Thus, as in the examples in the last section, to satisfy the initial condition we must be able to expand an arbitrary function in an infinite series of known functions, determined by a differential equation and a set of boundary conditions. However, although the functions in terms of which the expansion is to be carried out are sines, the values of λ appearing in their arguments are spaced at incommensurable intervals, and so the required series is *not* a Fourier series. Clearly, something is involved which includes Fourier series as a special case but is itself more general and more fundamental.

If we review thoughtfully our earlier discussion of Fourier series (Sec. 6.2), it should be apparent that the decisive property of the set of functions $\{\cos(n\pi x/l), \sin(n\pi x/l)\}$ which made it possible to determine one by one the coefficients in the assumed expansion

$$f(x) = \frac{1}{2}a_0 + a_1\cos\frac{\pi x}{l} + a_2\cos\frac{2\pi x}{l} + \cdots$$
$$+ b_1\sin\frac{\pi x}{l} + b_2\sin\frac{2\pi x}{l} + \cdots$$

was that the integral of the product of any two distinct members of the set taken over the appropriate interval is zero. For it was this that enabled us to multiply the series for $f(x)$ by $\cos n\pi x/l$ or $\sin n\pi x/l$ and eliminate all but one of the unknown coefficients simply by integrating from d to $d + 2l$.

Now, sines and cosines are by no means the only functions from which sets can be constructed having the property that the integral between suitable limits of the product of two distinct members of the set is zero. In fact, the trigonometric functions which appear in Fourier expansions are merely one of the simplest examples of infinitely many such systems of functions, whose existence we shall soon establish.

DEFINITION 1

If a sequence of real functions

$$\{\phi_n(x)\} \qquad n = 1, 2, 3, \ldots$$

which are defined over some interval (a,b), finite or infinite, has the property that

$$\int_a^b \phi_m(x)\phi_n(x)\,dx \begin{cases} = 0 & m \neq n \\ \neq 0 & m = n \end{cases}$$

then the functions are said to form an orthogonal set on that interval.

DEFINITION 2

If the functions of an orthogonal set $\{\phi_n(x)\}$ have the property that

$$\int_a^b \phi_n^2(x)\,dx = 1 \qquad \text{for all values of } n$$

then the functions are said to be orthonormal on the interval (a,b).

Any set of orthogonal functions can easily be converted into an orthonormal set. In fact, if the functions of the set $\{\phi_n(x)\}$ are orthogonal and if k_n is the (necessarily positive) value of $\int_a^b \phi_n^2(x)\,dx$, then the functions

$$\frac{\phi_1(x)}{\sqrt{k_1}}, \frac{\phi_2(x)}{\sqrt{k_2}}, \frac{\phi_3(x)}{\sqrt{k_3}}, \ldots$$

are clearly orthonormal. It is, therefore, no specialization to assume that an orthogonal set of functions is also orthonormal.

DEFINITION 3

If a sequence of real functions $\{\phi_n(x)\}$ has the property that, over some interval (a,b), finite or infinite,

$$\int_a^b p(x)\phi_m(x)\phi_n(x)\,dx \begin{cases} = 0 & m \neq n \\ \neq 0 & m = n \end{cases}$$

then the functions are said to be orthogonal with respect to the weight function $p(x)$ on that interval.

Any set of functions orthogonal with respect to a weight function $p(x)$ can be reduced to a system orthogonal in the first sense simply by multiplying each member of the set by $\sqrt{p(x)}$ if, as we shall suppose, $p(x) \geqq 0$ on the interval of orthogonality

With respect to any set of functions $\{\phi_n(x)\}$ orthogonal over an interval (a,b), an arbitrary function $f(x)$ has a formal expansion analogous to a Fourier expansion, for we can write

$$f(x) = a_1\phi_1(x) + a_2\phi_2(x) + \qquad + a_n\phi_n(x) + \tag{5}$$

Then, multiplying by $\phi_n(x)$ and formally integrating between the appropriate limits, a and b, we have

$$\int_a^b f(x)\phi_n(x)\,dx = a_1 \int_a^b \phi_1(x)\phi_n(x)\,dx$$
$$+ a_2 \int_a^b \phi_2(x)\phi_n(x)\,dx + \qquad + a_n \int_a^b \phi_n^2(x)\,dx + \quad \cdot$$

From the property of orthogonality, all integrals on the right are zero except the one which contains a square in its integrand Hence, we can solve at once for a_n as the quotient of two known integrals·

$$a_n = \frac{\int_a^b f(x)\phi_n(x)\,dx}{\int_a^b \phi_n^2(x)\,dx}$$

However, although the orthogonality of the ϕ's makes it possible to determine the coefficients in the expansion (5), this property is not sufficient to guarantee that this series converges to $f(x)$ or even converges at all.

To pursue this matter a little further, it is convenient to introduce the idea of a **null function**:

DEFINITION 4

A real function $f(x)$ is said to be a null function on the interval (a,b) if

$$\int_a^b f^2(x)\,dx = 0$$

If $f(x)$ is identically zero, it is obviously a null function However, a null function need not be identically zero In fact, since the area under a curve is not altered by changing the ordinate of the curve at one or more isolated points, it is clear that we can have $\int_a^b f^2(x)\,dx = 0$ even though $f(x)$ has nonzero values at a finite or countably infinite number of points between a and b On the other hand, if there is any subinterval of (a,b), no matter how short, at all points of which $f(x)$ is different from zero, then $\int_a^b f^2(x)\,dx \neq 0$ and $f(x)$ is not a null function From this it is not difficult to show that *a null function is zero at every point where it is continuous*

Clearly, any null function is orthogonal to every member of an orthogonal set $\{\phi_n(x)\}$ It is conceivable, also, that a nonnull function $f(x)$ might be orthogonal to every ϕ, that is, that we might have

$$\int_a^b f(x)\phi_n(x)\,dx = 0 \qquad \text{for all values of } n$$

In such a case, every coefficient in the expansion of $f(x)$ in terms of the ϕ's would be zero, and the series (5) would converge to zero at all points of (a,b) even though $f(x)$ was not a null function That this is actually possible is easily shown by example. For instance, although the functions $\{\sin nx\}$ are readily shown to be orthogonal over the interval $(-\pi,\pi)$, not every function can be represented on this interval by a series of the form

$$a_1 \sin x + a_2 \sin 2x + \qquad + a_n \sin nx +$$

In particular, if $f(x) = x^2$, we have, for the coefficients in its formal expansion,

$$a_n = \frac{\int_{-\pi}^{\pi} x^2 \sin nx\,dx}{\int_{-\pi}^{\pi} \sin^2 nx\,dx}$$

$$= \frac{1}{\pi}\left[\frac{2x}{n^2}\sin nx - \left(\frac{x^2}{n} - \frac{2}{n^3}\right)\cos nx\right]_{-\pi}^{\pi} = 0$$

for all values of n More generally, since every member of the set $\{\sin nx\}$ is odd, it is clear that no series of these functions can represent *any* even function on the interval $(-\pi,\pi)$

Evidently, important as it is, orthogonality is not the whole story, and the functions in our orthogonal systems must possess some further property before the expansion (5) can be used with confidence What is required is that the set of functions $\{\phi_n(x)\}$, in addition to being orthogonal, should also possess the property of **completeness** described in the following definition

DEFINITION 5

A set of orthogonal functions $\{\phi_n(x)\}$ is said to be complete if the relation $\int_a^b f(x)\phi_n(x)\,dx = 0$ can hold for all values of n only if $f(x)$ is a null function

If $\{\phi_n(x)\}$ is a complete orthogonal set, then clearly not all coefficients in the expansion of a nonnull function can be zero, and thus no nontrivial function can have a trivial expansion. In fact, we have the following theorem.

THEOREM 1

If the formal expansion

$$a_1\phi_1(x) + a_2\phi_2(x) + \cdots + a_n\phi_n(x) + \cdots$$

of a function $f(x)$ in terms of the members of a complete orthonormal set $\{\phi_n(x)\}$ converges and can be integrated term by term, then the sum of the series differs from $f(x)$ by at most a null function; that is, the sum of the series cannot differ from $f(x)$ over any interval of finite length

PROOF By hypothesis, the series $\sum_{n=1}^{\infty} a_n\phi_n(x)$ converges to some function, hence, it is meaningful to consider the difference

$$g(x) = f(x) - \sum_{n=1}^{\infty} a_n\phi_n(x)$$

If we can prove that $g(x)$ is a null function, the assertion of the theorem will be established To do this, consider

$$\begin{aligned}\int_a^b \phi_m(x)g(x)\,dx &= \int_a^b \phi_m(x)\left[f(x) - \sum_{n=1}^{\infty} a_n\phi_n(x)\right]dx \\ &= \int_a^b \phi_m(x)f(x)\,dx - \int_a^b \phi_m(x)\left[\sum_{n=1}^{\infty} a_n\phi_n(x)\right]dx \\ &= \int_a^b \phi_m(x)f(x)\,dx - \sum_{n=1}^{\infty} a_n\int_a^b \phi_m(x)\phi_n(x)\,dx \\ &= a_m - a_m \\ &\doteq 0 \qquad m = 1, 2, 3,\end{aligned}$$

Hence, $g(x)$ is orthogonal to every one of the ϕ's Therefore, since the ϕ's form a complete set, $g(x)$ must be a null function, and the theorem is established

Closely associated with the concept of completeness is the concept of **closure*** described in the following definitions

DEFINITION 6

If $\lim_{n\to\infty}\int_a^b [f(x) - S_n(x)]^2\,dx = 0$, the sequence of functions $S_n(x)$ is said to converge in the mean to $f(x)$

DEFINITION 7

If $S_n(x) = a_1\phi_1(x) + a_2\phi_2(x) + \cdots + a_n\phi_n(x)$ is the nth partial sum of the expansion of $f(x)$ in terms of the members of an orthonormal set $\{\phi_n(x)\}$ and if $S_n(x)$ converges in the mean to $f(x)$ for every $f(x)$, then the set $\{\phi_n(x)\}$ is said to be closed

One important property of closed orthonormal sets is contained in the so-called **theorem of Parseval:**

---

* What we have called *completeness* some authors call *closure*, and vice versa.

THEOREM 2

If $a_1\phi_1(x) + a_2\phi_2(x) + \qquad + a_n\phi_n(x) + \qquad$ is the expansion of a function $f(x)$ in terms of the members of a closed orthonormal set $\{\phi_n(x)\}$, then

$$\sum_{n=1}^{\infty} a_n^2 = \int_a^b f^2(x)\,dx$$

PROOF From the definition of closure, we have

$$\lim_{m\to\infty} \int_a^b \left[f(x) - \sum_{n=1}^{m} a_n\phi_n(x)\right]^2 dx = 0$$

or

$$\lim_{m\to\infty} \int_a^b \left[\{f(x)\}^2 - 2f(x)\sum_{n=1}^{m} a_n\phi_n(x) + \left\{\sum_{n=1}^{m} a_n\phi_n(x)\right\}^2\right] dx = 0$$

If we now perform the indicated integration, remembering that

$$\int_a^b f(x)\phi_n(x)\,dx = a_n$$

and observing that, in the integral of the last term

$$\int_a^b \phi_m(x)\phi_n(x)\,dx = \begin{cases} 0 & m \neq n \\ 1 & m = n \end{cases}$$

we obtain

$$\lim_{m\to\infty} \left\{\int_a^b [f(x)]^2\,dx - 2\sum_{n=1}^{m} a_n^2 + \sum_{n=1}^{m} a_n^2\right\} = 0$$

or

$$\sum_{n=1}^{\infty} a_n^2 = \int_a^b [f(x)]^2\,dx \qquad \text{as asserted}$$

As an immediate consequence of the last theorem, we have the following important result

THEOREM 3

A closed orthonormal system $\{\phi_n(x)\}$ is also complete

PROOF To prove this, let us suppose that the closed orthonormal system $\{\phi_n(x)\}$ is not complete This implies that there is at least one nonnull function $f(x)$ which is orthogonal to each of the ϕ's and which, therefore, has the property that every coefficient in its expansion in terms of the ϕ's is zero However, since the set $\{\phi_n(x)\}$ is closed, we have, from Parseval's theorem,

$$\int_a^b f^2(x)\,dx = \sum_{n=1}^{\infty} a_n^2$$

Hence, since each a_n is zero, as we have just observed, it follows that $f(x)$ is a null function, contrary to our assumption. This contradiction forces us to abandon the supposition that the closed set $\{\phi_n(x)\}$ is incomplete, and the theorem is established

The converse of Theorem 3 is also true, but the proof of this fact is difficult, and we shall not attempt it

A great deal of important advanced mathematics deals with the properties of special orthogonal systems and with the validity of the formal expansion we have just created. In the next chapter we shall examine in some detail two such systems, namely, the Bessel functions and the Legendre polynomials Questions concerning the convergence of the generalized Fourier series (5), however, we shall not discuss, and in our work we shall assume not only that all the expansions we obtain converge but also that they actually represent the functions which generated them

Orthogonal functions arise naturally and inevitably in many types of problems in pure and applied mathematics [*] Their existence in problems such as we have been considering is guaranteed by the following beautiful and important theorem :[†]

THEOREM 4

Given the differential equation

$$\frac{d[r(x)y']}{dx} + [q(x) + \lambda p(x)]y = 0$$

where $r(x)$ and $p(x)$ are continuous on the closed interval $a \leqq x \leqq b$ and $q(x)$ is continuous at least over the open interval $a < x < b$ If $\lambda_1, \lambda_2, \lambda_3, \ldots$ are the values of the parameter λ for which there exist solutions of this equation possessing continuous first derivatives and satisfying the boundary conditions

$$a_1y(a) - a_2y'(a) = 0$$
$$b_1y(b) - b_2y'(b) = 0$$

where a_1, a_2, b_1, b_2 are any constants such that a_1 and a_2 are not both zero and b_1 and b_2 are not both zero, and if $y_1, y_2, y_3, \ldots$ are the solutions corresponding to these values of λ, then the functions $\{y_n(x)\}$ form a system orthogonal with respect to the weight function $p(x)$ over the interval (a,b)

PROOF To prove this, let y_m and y_n be the solutions associated with two distinct values of λ, say λ_m and λ_n This means that

$$\frac{d(ry'_m)}{dx} + (q + \lambda_m p)y_m = 0$$
$$\frac{d(ry'_n)}{dx} + (q + \lambda_n p)y_n = 0$$

Now multiply the first of these equations by y_n and the second by y_m and then subtract the second equation from the first. The result, after transposing, is

$$(\lambda_m - \lambda_n)py_my_n = y_m\frac{d(ry'_n)}{dx} - y_n\frac{d(ry'_m)}{dx}$$

or, integrating between a and b,

$$(\lambda_m - \lambda_n)\int_a^b py_my_n\,dx = \int_a^b \left[y_m\frac{d(ry'_n)}{dx}\right]dx - \int_a^b \left[y_n\frac{d(ry'_m)}{dx}\right]dx$$

---

[*] It is interesting and instructive in this connection to reread the discussion of orthogonal polynomials in Sec 4 6 and to refer to the discussion of the orthogonality of vectors in Sec 10 4

[†] This theorem and the boundary value problem with which it deals are usually associated with the names of the Swiss mathematician J. C F. Sturm (1803-1855) and the French mathematician Joseph Liouville (1809-1882

If we can prove that the integral on the left vanishes whenever m and n are different, we shall have established the orthogonality property of the functions of the set $\{y_n(x)\}$. This we shall prove by showing that the right-hand side of the last equation is zero To do this we begin by integrating the terms on the right by parts

$$\int_a^b \left[y_m \frac{d(ry_n')}{dx} \right] dx \xrightarrow[\substack{du = y_m'\,dx \qquad v = ry_n'}]{\substack{u = y_m \qquad dv = d(ry_n')}} ry_my_n' \Big|_a^b - \int_a^b ry_n'y_m'\,dx$$

$$\int_a^b \left[y_n \frac{d(ry_m')}{dx} \right] dx \xrightarrow[\substack{du = y_n'\,dx \qquad v = ry_m'}]{\substack{u = y_n \qquad dv = d(ry_m')}} ry_ny_m' \Big|_a^b - \int_a^b ry_m'y_n'\,dx$$

When we subtract these expressions, the integrals which remain on the right cancel, and we have

$$\int_a^b \left[y_m \frac{d(ry_n')}{dx} \right] dx - \int_a^b \left[y_n \frac{d(ry_m')}{dx} \right] dx = r(y_my_n' - y_m'y_n) \Big|_a^b \tag{6}$$

Now y_m and y_n are not merely solutions of the given differential equation For every m and n, they also satisfy the boundary conditions

$$a_1y(a) = a_2y'(a) \qquad \text{and} \qquad b_1y(b) = b_2y'(b)$$

Substituting for $y'(a)$ and $y'(b)$ from these expressions into the evaluated antiderivative in (6), we obtain

$$\begin{aligned} \int_a^b \left[y_m \frac{d(ry_n')}{dx} \right] dx &- \int_a^b \left[y_n \frac{d(ry_m')}{dx} \right] dx \\ &= r(b) \left[y_m(b) \frac{b_1}{b_2} y_n(b) - \frac{b_1}{b_2} y_m(b)y_n(b) \right] \\ &\quad - r(a) \left[y_m(a) \frac{a_1}{a_2} y_n(a) - \frac{a_1}{a_2} y_m(a)y_n(a) \right] \equiv 0 \end{aligned}$$

If a_2 or b_2, or both, should be zero, this result can still be established by substituting for $y(a)$ or $y(b)$, or both, instead of for their derivatives, since a_1 and a_2 cannot both vanish nor can b_1 and b_2 Moreover, if $r(a) = 0$, then the first boundary condition becomes irrelevant, that is, the integrated terms vanish at $x = a$ without the need of any condition on the solutions y_m and y_n Likewise, if $r(b) = 0$, the second boundary condition is irrelevant We have thus shown that under the conditions of the theorem

$$(\lambda_m - \lambda_n) \int_a^b py_my_n\,dx = 0$$

Since λ_m and λ_n were any two *distinct* values of λ, the difference $\lambda_m - \lambda_n$ cannot vanish Hence

$$\int_a^b py_my_n\,dx = 0$$

and the theorem is established

In each of the torsional vibration problems we considered in Sec 8 4, the functions in terms of which we had to expand the initial conditions satisfied a differential equation and a set of boundary conditions included under Theorem 4. This, and not the coincidental fact that Fourier series were involved, explains why the final expansion could be carried out in each case

EXAMPLE 1 *(continued)*

When we left Example 1 in order to develop the theory necessary to complete its solution, we were faced with the necessity of expanding the initial temperature $u(x,0) = f(x)$ in a series of the form (4),

$$f(x) = \sum_{n=1}^{\infty} B_n \sin \lambda_n x$$

where the functions $\{\sin \lambda_n x\}$ were the solutions of the differential equation

$$X'' + \lambda^2 X = 0$$

which satisfied the conditions

$$X(0) = 0$$

$$hX(l) + X'(l) = 0$$

But this equation and the accompanying boundary conditions are in all respects a special case covered by Theorem 4. In fact, with λ^2 written in place of λ, we have

$$r(x) = 1 \qquad q(x) = 0 \qquad p(x) = 1$$

$$a = 0 \qquad b = l$$

$$a_1 = 1 \qquad a_2 = 0 \qquad b_1 = h \qquad b_2 = -1$$

Hence, by the last theorem, the functions $\{\sin \lambda_n x\}$ form a set orthogonal with respect to the weight function $p(x) = 1$ on the interval $(0,l)$.

To determine B_n we now multiply Eq. (4) by $\sin \lambda_n x$ and formally integrate from 0 to l. Because of the orthogonality of the functions $\sin \lambda_n x$, every integral on the right vanishes except the one whose integrand contains $\sin^2 \lambda_n x$. Therefore,

$$B_n = \frac{\int_0^l f(x) \sin \lambda_n x \, dx}{\int_0^l \sin^2 \lambda_n x \, dx}$$

or, evaluating the integral in the denominator and recalling that $z_n = \lambda_n l$ satisfies the equation $\sin z_n = -\alpha z_n \cos z_n$,

$$B_n = \frac{2}{l(1 + \alpha \cos^2 z_n)} \int_0^l f(x) \sin \lambda_n x \, dx$$

With B_n determined, the formal solution is now complete.

Problems involving second-order differential equations are not the only ones in which orthogonal functions arise. In particular, we have the following important theorem covering fourth-order systems, of which the vibrating beam is a special case.

THEOREM 5

Given the differential equation

$$\frac{d^2[r(x)y'']}{dx^2} + [q(x) + \lambda p(x)]y = 0$$

where $r(x)$ and $p(x)$ are continuous on the closed interval (a,b) and $q(x)$ is continuous at least on the open interval (a,b). If $\lambda_1, \lambda_2, \lambda_3, \ldots$ are the values of the parameter λ for which there exist solutions of this equation possessing con tinuous third derivatives and satisfying the boundary conditions

$$a_1y(a) - \alpha_1(ry'')'\Big|_a = 0 \qquad a_2y'(a) - \alpha_2(ry'')\Big|_a = 0$$

$$b_1y(b) - \beta_1(ry'')'\Big|_b = 0 \qquad b_2y'(b) - \beta_2(ry'')\Big|_b = 0$$

where neither a_i and α_i nor b_i and β_i are both zero, and if $y_1, y_2, y_3,$ are the solutions corresponding to these values of λ, then the functions $\{y_n(x)\}$ form a system orthogonal with respect to the weight function $p(x)$ over the interval (a,b)

EXAMPLE 2

A uniform cantilever of length l begins to vibrate with initial displacement $y(x,0) = f(x)$ and initial velocity $\dfrac{\partial y}{\partial t}\Big|_{x,0} = g(x)$ Find its displacement at any point at any subsequent time

For definiteness let us assume that the built-in end of the beam is at the origin Then, since the beam is of uniform cross section and bears no external load, we have to solve Eq (11), Sec 8 2,

$$a^2\frac{\partial^4 y}{\partial x^4} = -\frac{\partial^2 y}{\partial t^2}$$

subject to the boundary conditions

$y(0,t) = 0$ i e , the displacement at the built-in end is zero

$\dfrac{\partial y}{\partial x}\Big|_{0,t} = 0$ i e , the slope at the built in end is zero

$\dfrac{\partial^2 y}{\partial x^2}\Big|_{l,t} = 0$ i.e , the moment $EI\dfrac{\partial^2 y}{\partial x^2}$ at the free end is zero

$\dfrac{\partial^3 y}{\partial x^3}\Big|_{l,t} = 0$ i e , the shear $\dfrac{\partial\left(EI\dfrac{\partial^2 y}{\partial x^2}\right)}{\partial x}$ at the free end is zero

As usual, we begin by assuming a product solution $y(x,t) = X(x)T(t)$, substituting it into the given partial differential equation, getting $a^2X^{IV}T = -XT''$, and then separating variables,

$$a^2\frac{X^{IV}}{X} = -\frac{T''}{T}$$

Since x and t are independent variables, these two fractions must have a common constant value, say μ If $\mu \leqq 0$, the solution for T cannot be periodic, as we know it must be to represent undamped vibrations Hence we restrict μ to be positive,* and write $\mu = \lambda^2$ This leads to the component differential equations

$$T'' = -\lambda^2 T \qquad \text{and} \qquad X^{IV} = \frac{\lambda^2}{a^2}X$$

and the respective solutions

$$T = A\cos\lambda t + B\sin\lambda t \tag{7}$$

$$X = C\cos\sqrt{\frac{\lambda}{a}}\,x + D\sin\sqrt{\frac{\lambda}{a}}\,x + E\cosh\sqrt{\frac{\lambda}{a}}\,x + F\sinh\sqrt{\frac{\lambda}{a}}\,x \tag{8}$$

---

* In vibration problems where it is clear that only periodic solutions are possible, engineers often take their initial assumption to be

$$y(x,t) = X(x)(A\cos\lambda t + B\sin\lambda t)$$

as, in effect, we did in Example 3, Sec 5.5, in studying the undamped vibrations of an electrical network with only a finite number of degrees of freedom.

Imposing the first boundary condition, namely,

$$y(0,t) \equiv X(0)T(t) = 0$$

we find

$$(C + E)T(t) = 0$$

Since we cannot permit $T(t)$ to be identically zero without having the entire solution become trivial, we conclude that

$$C + E = 0$$

Imposing the second boundary condition, namely,

$$\left.\frac{\partial y}{\partial x}\right|_{0,t} \equiv X'(0)T(t) = 0$$

we find $(D + F)\sqrt{\dfrac{\lambda}{a}}\,T(t) = 0$ Hence, $D + F = 0$

From the third and fourth boundary conditions

$$\left.\frac{\partial^2 y}{\partial x^2}\right|_{l,t} \equiv X''(l)T(t) = 0 \qquad \text{and} \qquad \left.\frac{\partial^3 y}{\partial x^3}\right|_{l,t} \equiv X'''(l)T(t) = 0$$

we obtain, respectively,

$$\left(-C\cos\sqrt{\frac{\lambda}{a}}\,l - D\sin\sqrt{\frac{\lambda}{a}}\,l + E\cosh\sqrt{\frac{\lambda}{a}}\,l + F\sinh\sqrt{\frac{\lambda}{a}}\,l\right)\frac{\lambda}{a}\,T(t) = 0$$

and

$$\left(C\sin\sqrt{\frac{\lambda}{a}}\,l - D\cos\sqrt{\frac{\lambda}{a}}\,l + E\sinh\sqrt{\frac{\lambda}{a}}\,l + F\cosh\sqrt{\frac{\lambda}{a}}\,l\right)\left(\frac{\lambda}{a}\right)^{3/2}T(t) = 0$$

Hence, for convenience, setting

$$(9) \qquad z = \sqrt{\frac{\lambda}{a}}\,l$$

we must have

$$-C\cos z - D\sin z + E\cosh z + F\sinh z = 0$$

$$C\sin z - D\cos z + E\sinh z + F\cosh z = 0$$

If we eliminate C and D from these equations by using the conditions

$$C + E = 0 \qquad \text{and} \qquad D + F = 0$$

we obtain the system

$$(10) \qquad E(\cosh z + \cos z) + F(\sinh z + \sin z) = 0$$

$$E(\sinh z - \sin z) + F(\cosh z + \cos z) = 0$$

These equations will have a nontrivial solution for E and F if and only if the determinant of the coefficients is equal to zero Hence we must have

$$\begin{vmatrix} \cosh z + \cos z & \sinh z + \sin z \\ \sinh z - \sin z & \cosh z + \cos z \end{vmatrix} = 0$$

or, expanding and simplifying,

$$\cosh z \cos z = -1$$

The existence of infinitely many roots of this equation, i.e., $\cos z = -1/\cosh z$, can be inferred from Fig 8 13, where the graphs of

$$y = \cos z \qquad \text{and} \qquad y = \frac{-1}{\cosh z}$$

are plotted.

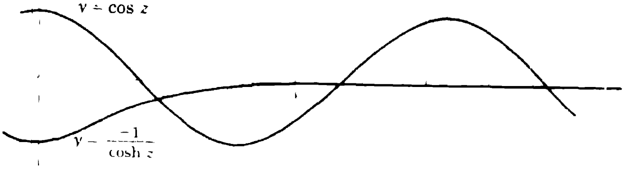


FIGURE 8.13 Plot showing the graphical solution of the equation $\cos z = -1/\cosh z$.

From these roots $z_1, z_2, z_3, \ldots$, we can find the relevant values of λ at once from Eq. (9)

$$\lambda_1 = \frac{az_1^2}{l^2}, \qquad \lambda_2 = \frac{az_2^2}{l^2}, \qquad \lambda_3 = \frac{az_3^2}{l^2}, \qquad \ldots$$

When z has any one of the values $z_1, z_2, z_3, \ldots$, the equations (10) become dependent, and we can write either

$$\frac{E}{F} = -\frac{\sinh z + \sin z}{\cosh z + \cos z} \qquad or \qquad \frac{E}{F} = -\frac{\cosh z + \cos z}{\sinh z - \sin z}$$

as we choose. Using the former, we have

$$E_n = -C_n = -(\sinh z_n + \sin z_n)K_n$$
$$F_n = -D_n = (\cosh z_n + \cos z_n)K_n$$

where K_n is an arbitrary constant. Therefore, substituting into Eq. (8),

$$\begin{aligned} X_n(x) &= C_n \cos\sqrt{\frac{\lambda_n}{a}}\,x + D_n \sin\sqrt{\frac{\lambda_n}{a}}\,x + E_n \cosh\sqrt{\frac{\lambda_n}{a}}\,x + F_n \sinh\sqrt{\frac{\lambda_n}{a}}\,x \\ &= K_n(\sinh z_n + \sin z_n)\left(\cos z_n\frac{x}{l} - \cosh z_n\frac{x}{l}\right) \\ &\qquad - K_n(\cosh z_n + \cos z_n)\left(\sin z_n\frac{x}{l} - \sinh z_n\frac{x}{l}\right) \end{aligned}$$

Hence, absorbing K_n in the coefficients A_n and B_n in the expression (7) for T_n and redefining X_n to be the completely determined function

(11) $$\begin{aligned} X_n(x) &= (\sinh z_n + \sin z_n)\left(\cos z_n\frac{x}{l} - \cosh z_n\frac{x}{l}\right) \\ &\qquad - (\cosh z_n + \cos z_n)\left(\sin z_n\frac{x}{l} - \sinh z_n\frac{x}{l}\right) \end{aligned}$$

we have, as the formal solution of the partial differential equation which meets the four given boundary conditions,

$$y(x,t) = \sum_{n=1}^{\infty} X_n(x)T_n(t) = \sum_{n=1}^{\infty} X_n(x)(A_n \cos \lambda_n t + B_n \sin \lambda_n t)$$

To satisfy the initial displacement condition, we must have

(12) $$y(x,0) \equiv f(x) = \sum_{n=1}^{\infty} A_n X_n(x)$$

and to satisfy the initial velocity condition, we must have

(13) $$\left.\frac{\partial y}{\partial t}\right|_{x,0} \equiv g(x) = \sum_{n=1}^{\infty} (\lambda_n B_n) X_n(x)$$

Thus, again, to satisfy the initial conditions we must be able to expand an arbitrary function in an infinite series of known functions, in this case the functions of the set $\{X_n(x)\}$ defined by Eq. (11). These bear little or no resemblance to the terms of a Fourier series, but the required expansions can easily be carried out using the orthogonality of the X_n's, which is guaranteed by Theorem 5 (and of course their completeness, which as usual we must assume). In fact, with λ^2/a^2 written in place of λ, our problem is just the special case of Theorem 5 for which

$$\begin{array}{llll} r(x) = 1 & q(x) = 0 & p(x) = 1 & \\ a = 0 & b = l & & \\ a_1 = 1 & \alpha_1 = 0 & b_1 = 0 & \beta_1 = -1 \\ a_2 = 1 & \alpha_2 = 0 & b_2 = 0 & \beta_2 = -1 \end{array}$$

Hence, the functions of the set $\{X_n(x)\}$ are orthogonal with respect to the weight function $p(x) = 1$ over the interval $(0,l)$.

With the orthogonality of the X_n's now established, we can determine A_n and B_n immediately by multiplying Eqs. (12) and (13) by $X_n(x)$ and integrating from 0 to l. The results are

$$A_n = \frac{\int_0^l f(x)X_n(x)\,dx}{\int_0^l X_n^2(x)\,dx} \qquad \text{and} \qquad B_n = \frac{\int_0^l g(x)X_n(x)\,dx}{\lambda_n \int_0^l X_n^2(x)\,dx}$$

We are now in a position to summarize the main features of a simple boundary value problem. By assuming that solutions for the dependent variable exist in the form of products of functions of the respective independent variables, the original differential equation is broken down into several ordinary differential equations, each of which involves a parameter λ which ranges over a continuous infinity of values.

When the boundary conditions of the problem are imposed upon the product solutions obtained by solving the component ordinary differential equations, it is necessary, in order to avoid solutions which are identically zero, that the parameter λ satisfy a certain equation. This equation is known as the **characteristic equation** of the problem, and its roots, in general infinite in number, are known as the **characteristic values** or **eigenvalues** or **Eigenwerte*** of the problem. Only for them can solutions be found satisfying both the partial differential equation and the given boundary conditions. In a vibration problem, the characteristic values determine the natural frequencies of the system, and the characteristic equation is therefore usually called the **frequency equation.** The solutions which correspond to the respective characteristic values are known as the **characteristic functions** or **eigenfunctions** of the problem. In a vibration problem, they are usually called the **normal modes,** since they define the relative amplitudes of the extreme positions between which the system oscillates when it is vibrating at a single natural frequency, i.e., in a "normal" manner.

To satisfy the initial conditions of the problem it is necessary to be able to express an arbitrary function as an infinite series of the characteristic functions of the problem This can be done in most cases of interest because under very general conditions the characteristic functions of a boundary value problem form an orthogonal set over the particular interval related to the problem.

EXERCISES

1 If $r(a) = r(b)$, show that the conclusion of Theorem 4 follows equally well if the boundary conditions are of the form

$$y(a) = y(b) \qquad \text{and} \qquad y'(a) = y'(b)$$

2 Verify by direct integration that the characteristic functions of Example 1 are orthogonal over the interval $(0,l)$

3 In Example 1, if $\alpha = 1$, compute the values of z_1, z_2, and z_3, and determine the first three coefficients in the expansion of the initial condition $u(x,0) = f(x) = x$

4 In Example 1, if the left end of the rod is perfectly insulated, determine the characteristic equation, show that it has infinitely many roots, and prove that $z_{n+1} - z_n$ approaches π as n becomes infinite

5 Find the temperature $u(x,t)$ in a slender rod of length l whose curved surface and left end are perfectly insulated and whose right end radiates freely into air of constant temperature 0° if the rod is initially at the temperature 100° throughout

6 In Example 1, if both ends of the rod radiate freely into air of constant temperature 0°, determine the characteristic equation, show that it has infinitely many roots, and prove that $z_{n+1} - z_n$ approaches π as n becomes infinite

7 A slender rod of length l has its curved surface and left end perfectly insulated Its right end radiates freely into air of constant temperature 70° Initially the temperature throughout the rod is 100° Find the temperature of the rod as a function of x and t (Hint Let $U = u - 70$ be the dependent variable)

8 Work Example 1 with the left end of the rod maintained at the constant temperature 100°

9 Prove Theorem 5

10 Prove that the general linear second-order differential equation

$$p_0(x)y'' + p_1(x)y' + p_2(x)y = \lambda y$$

can be reduced to an equation of the Sturm-Liouville form (see Theorem 4) by multiplying it by the factor

$$\frac{1}{p_0(x)} e^{\int_{x_0}^{x} [p_1(x)/p_0(x)]dx}$$

11 Find the frequency equation and the normal modes for the transverse vibration of a uniform beam whose ends are

a Hinged-hinged (Hint A **hinged end** is one where a beam, though constrained so it cannot deflect, is still free to turn, i e , an end where both the displacement and the moment are zero at all times A hinged end is often referred to as a **simply supported end.**)

b Fixed-fixed

c Free-free

d Fixed-hinged

e Free-hinged

12 Find the frequency equation for the transverse vibrations of a uniform cantilever bearing a concentrated mass at the free end (Hint At the free end one boundary condition is that the shear, instead of being zero, is equal to the inertia force of the attached mass)

13 Find the frequency equation for the transverse vibrations of a uniform hinged-hinged beam bearing a concentrated mass at its mid-point

14 Find the frequency equation for a uniform torsional cantilever if a disk of polar moment of

inertia I_p is attached to the free end of the shaft. (Hint: At the free end of the shaft the boundary condition is that the torque, instead of being zero, is equal to the inertia torque of the disk.)

15 Show that the normal modes in Exercise 14 are not orthogonal.

16 Find the frequency equation for the torsional vibrations of a shaft of length $2l$ which is clamped at $x = 0$ and free at $x = 2l$ if the radius of the portion of the shaft between $x = 0$ and $x = l$ is r_1 and the radius of the portion of the shaft between $x = l$ and $x = 2l$ is r_2. [Hint: Solve the problem separately for the interval $(0,l)$ and the interval $(l,2l)$ and then, in addition to the two end conditions, impose the conditions that at $x = l$ both the angle of twist and the transmitted torque are continuous.]

17 Show that the system $\{\cos nx\}$, $n = 0, 1, 2, 3, \ldots$, is orthogonal but not complete over the interval $(-\pi,\pi)$.

18 Show that for an orthonormal system $\{\phi_n(x)\}$, whether closed or not, we have **Bessel's inequality**

$$\sum_{n=1}^{\infty} a_n^2 \leqq \int_a^b [f(x)]^2\,dx$$

where the a's are the coefficients in the generalized Fourier expansion of $f(x)$ in terms of the ϕ's and (a,b) is the interval of orthogonality. Using this result, show that

$$\lim_{n\to\infty} \int_a^b \phi_n(x)f(x)\,dx = 0$$

19 If $\{\phi_n(x)\}$ is an orthonormal set over the interval (a,b), show that the values of the c's which make

$$\int_a^b [f(x) - c_1\phi_1(x) - c_2\phi_2(x) - \cdots - c_n\phi_n(x)]^2\,dx$$

a minimum are the corresponding coefficients in the generalized Fourier expansion of $f(x)$ in terms of the ϕ's.

20 What is the minimum value of the integral in Exercise 19?

8.6

Further applications

Many problems in partial differential equations involve features not found in the simple examples we have used to elaborate the standard, elementary theory. We cannot here investigate in detail the variations and extensions of this theory, but, as illustrations, we shall present several additional examples exhibiting techniques of practical interest. In the first example, we shall see how the analysis of the forced vibrations of a continuous system leads to a nonhomogeneous rather than a homogeneous partial differential equation. In the second, we shall see how Fourier integrals, rather than Fourier series, enter into problems where the boundary conditions fail to provide a characteristic equation and λ remains a continuous parameter. In the third, we shall see that, though a partial differential equation may be separable, it may be impossible to make its product solutions fit the boundary conditions and so other methods must be used to solve it. In the fourth, we shall see how a partial differential equation involving three rather than two independent variables leads to a *double* series of characteristic functions and *two* separate expansion problems The important matter of the application of Laplace transform methods to the solution of partial differential equations we shall consider in the next section

EXAMPLE 1

A uniform string of length l is acted upon by a distributed periodic force

$$f(x,t) = \frac{w}{g}\phi(x)\sin\omega t$$

If the string is initially at rest in its equilibrium position, determine its subsequent motion given that frictional effects are negligible

From Eq (2), Sec 8 2, it is clear that the deflection of the string satisfies the partial differential equation

$$(1)\qquad \frac{\partial^2 y}{\partial t^2} = a^2\frac{\partial^2 y}{\partial x^2} + \phi(x)\sin\omega t$$

As in the case of a system with a single degree of freedom, the motion of the string consists of two parts, one described by the solution of the homogeneous equation

$$(2)\qquad \frac{\partial^2 y}{\partial t^2} = a^2\frac{\partial^2 y}{\partial x^2}$$

and the other described a particular solution corresponding to the nonhomogeneous term $\phi(x)\sin\omega t$

To find the solution of the homogeneous equation (2), that is, to determine the free motion of the string, we assume a product solution

$$y_H(x,t) = X(x)T(t)$$

and proceed *exactly* as we did in solving the wave equation for the torsional vibrations of a fixed-fixed shaft of uniform cross section in Sec 8 4 The result is

$$(3)\qquad y_H(x,t) = \sum_{n=1}^{\infty}\sin\frac{n\pi x}{l}\left(A_n\cos\frac{n\pi at}{l} + B_n\sin\frac{n\pi at}{l}\right)$$

To find a particular solution of the nonhomogeneous equation (1), we observe that, from physical considerations, the motion produced by the applied force must be periodic with the same period as the force Moreover, since the system is assumed to be frictionless, it is clear that the motion of the string must be in phase with the force Hence it is reasonable to assume a solution of the form

$$Y(x,t) = \Phi(x)\sin\omega t$$

We can now proceed in either of two ways In the first place, we can substitute $Y(x,t)$ into the nonhomogeneous equation (1), divide out the common factor $\sin\omega t$, solve the resulting nonhomogeneous ordinary differential equation namely,

$$-\omega^2\Phi = a^2\Phi'' + \phi(x)$$

and impose upon it the boundary conditions that

$$Y(0,t) = Y(l,t) = 0$$

i e , the conditions that

$$\Phi(0) = \Phi(l) = 0$$

When $\Phi(x)$ has been determined so that these conditions are fulfilled, we can then construct the complete solution

$$(4)\qquad y(x,t) = y_H(x,t) + Y(x,t)$$

$$= \sum_{n=1}^{\infty} \sin\frac{n\pi x}{l}\left(A_n\cos\frac{n\pi at}{l} + B_n\sin\frac{n\pi at}{l}\right) + \Phi(x)\sin\omega t$$

The initial conditions can now be imposed, giving

$$(5)\qquad y(x,0) = 0 = \sum_{n=1}^{\infty} A_n\sin\frac{n\pi x}{l}$$

and

$$(6)\qquad y(x,0) = 0 = \sum_{n=1}^{\infty}\frac{n\pi a}{l}B_n\sin\frac{n\pi x}{l} + \omega\Phi(x)$$

From (5) we conclude that the A's are the coefficients in the half-range sine expansion of 0, hence, $A_n = 0$ for all values of n. From (6) we conclude, similarly, that the terms $\left\{\frac{n\pi a}{l}B_n\right\}$ are the coefficients in the half-range sine expansion of $-\omega\Phi(x)$, hence,

$$B_n = -\frac{2\omega}{n\pi a}\int_0^l \Phi(x)\sin\frac{n\pi x}{l}\,dx$$

provided that ω is not equal to one of the natural frequencies $\{\omega_n\} = \{n\pi a/l\}$ of the system, i.e., provided that the system is not being "driven" at resonance.

On the other hand, before substituting $Y(x,t) = \Phi(x)\sin\omega t$ into the nonhomogeneous equation (1), we can expand $\phi(x)$ into a half-range sine series, getting, say,

$$\phi(x) = \sum_{n=1}^{\infty} C_n\sin\frac{n\pi x}{l}\qquad\text{where}\quad C_n = \frac{2}{l}\int_0^l \phi(x)\sin\frac{n\pi x}{l}\,dx$$

Then, assuming for $\Phi(x)$ a half-range sine expansion with undetermined coefficients, say

$$\Phi(x) = \sum_{n=1}^{\infty} D_n\sin\frac{n\pi x}{l}$$

we have, on substituting

$$Y(x,t) = \left(\sum_{n=1}^{\infty} D_n\sin\frac{n\pi x}{l}\right)\sin\omega t$$

into Eq. (1) and then dividing out $\sin\omega t$,

$$-\omega^2\sum_{n=1}^{\infty} D_n\sin\frac{n\pi x}{l} = -a^2\sum_{n=1}^{\infty} D_n\left(\frac{n\pi}{l}\right)^2\sin\frac{n\pi x}{l} + \sum_{n=1}^{\infty} C_n\sin\frac{n\pi x}{l}$$

Making this relation an identity by equating to zero the coefficient of $\sin n\pi x/l$ for each n, we find

$$D_n = \frac{C_n}{(n\pi a/l)^2 - \omega^2} = \frac{C_n}{\omega_n^2 - \omega^2}$$

Hence,

$$\Phi(x) = \sum_{n=1}^{\infty}\frac{C_n}{\omega_n^2 - \omega^2}\sin\frac{n\pi x}{l}$$

$$Y(x,t) = \Phi(x)\sin\omega t = \left(\sum_{n=1}^{\infty}\frac{C_n}{\omega_n^2 - \omega^2}\sin\frac{n\pi x}{l}\right)\sin\omega t$$

and the complete formal solution of the nonhomogeneous equation becomes

$$y(x,t) = \sum_{n=1}^{\infty} \sin\frac{n\pi x}{l}\left(A_n \cos\frac{n\pi at}{l} + B_n \sin\frac{n\pi at}{l}\right) + \left(\sum_{n=1}^{\infty} \frac{C_n}{\omega_n^2 - \omega^2}\sin\frac{n\pi x}{l}\right)\sin\omega t$$

$$= \sum_{n=1}^{\infty} \sin\frac{n\pi x}{l}\left(A_n \cos\frac{n\pi at}{l} + B_n \sin\frac{n\pi at}{l} + \frac{C_n}{\omega_n^2 - \omega^2}\sin\omega t\right)$$

To satisfy the initial conditions, we must have

$$y(x,0) = 0 = \sum_{n=1}^{\infty} A_n \sin\frac{n\pi x}{l}$$

whence, $A_n = 0$, and

$$y(x,0) = 0 = \sum_{n=1}^{\infty} \sin\frac{n\pi x}{l}\left(\frac{n\pi a}{l}B_n + \frac{\omega C_n}{\omega_r^2 - \omega^2}\right)$$

whence,
$$\frac{n\pi a}{l}B_n + \frac{\omega C_n}{\omega_n^2 - \omega^2} = 0 \qquad \text{or} \qquad B_n = \frac{\omega l C_n}{n\pi a(\omega^2 - \omega_n^2)}$$

From the expression for B_n it is clear that the frequency ω of the impressed force must not coincide with any natural frequency ω_n of the string unless the corresponding coefficient C_n is equal to zero, that is, unless the term

$$\sin\frac{n\pi x}{l} \equiv \sin\omega_n\frac{x}{a}$$

is missing from the half-range sine expansion of $\phi(x)$ If $\omega = \omega_n$ and $C_n \neq 0$ for some particular value of n, then the string is effectively being driven at a condition of resonance, and displacements of arbitrarily large amplitudes will be built up (See Exercise 1)

EXAMPLE 2

A slender rod whose curved surface is perfectly insulated stretches from $x = 0$ to $x = \infty$ Find the temperature in the rod as a function of x and t if the left end of the rod is maintained at the constant temperature 0° and if initially the temperature along the rod is given by $u(x,0) = f(x)$

Exactly as in Example 1, Sec 8 5, we find that the function

$$u = Be^{-\lambda^2 t/a^2}\sin\lambda x$$

satisfies the heat equation

$$\frac{\partial^2 u}{\partial x^2} = a^2\frac{\partial u}{\partial t}$$

and the boundary condition at the left end of the rod,

$$u(0,t) = 0$$

Lacking a second boundary condition, however, we have no further restriction on λ Therefore, instead of having an infinite set of *discrete* characteristic values λ_n, with corresponding solutions

$$u_n(x,t) = B_n e^{-\lambda_n^2 t/a^2}\sin\lambda_n x$$

we have a continuous family of solutions

$$u_\lambda(x,t) = B(\lambda)e^{-\lambda^2 t/a^2}\sin\lambda x$$

where the arbitrary constant B is now associated not with n, but with the continuous parameter λ, which can assume *any* real value

We cannot speak of an infinite series of particular solutions in this case Instead of *adding* the product solutions for each value of n we therefore try *integrating* them over all values of λ,

$$u(x,t) = \int_{-\infty}^{\infty} B(\lambda) e^{-\lambda^2 t/a^2} \sin \lambda x \, d\lambda \tag{7}$$

By direct substitution it is easily verified that this integral is a solution of the heat equation

It is now necessary to impose the initial condition $u(x,0) = f(x)$ on the solution $u(x,t)$ Setting $t = 0$ in Eq (7), it is clear that this requires that

$$f(x) = \int_{-\infty}^{\infty} B(\lambda) \sin \lambda x \, d\lambda$$

But this is just an instance of the Fourier integral we considered in Sec 6 7 There, in discussing what we called Fourier sine integrals, we saw [Eq (15a), Sec 6 7] that if

$$f(x) = \int_{-\infty}^{\infty} B(\lambda) \sin \lambda x \, d\lambda$$

then the coefficient function $B(\lambda)$ is given by

$$B(\lambda) = \frac{1}{\pi} \int_0^{\infty} f(x) \sin \lambda x \, dx$$

Introducing the dummy variable s for x in the integral defining $B(\lambda)$, we can, therefore, write Eq (7) in the form

$$\begin{aligned} u(x,t) &= \int_{-\infty}^{\infty} e^{-\lambda^2 t/a^2} \left[\frac{1}{\pi} \int_0^{\infty} f(s) \sin \lambda s \, ds \right] \sin \lambda x \, d\lambda \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_0^{\infty} e^{-\lambda^2 t/a^2} f(s) \sin \lambda s \sin \lambda x \, ds \, d\lambda \end{aligned}$$

which is the required solution

EXAMPLE 3

Find the steady-state potential at any point of an infinitely long transmission line if a signal voltage $E_0 \cos \omega t$ is applied at the sending end $x = 0$

Here we have to solve the telephone equation

$$\frac{\partial^2 e}{\partial x^2} = LC \frac{\partial^2 e}{\partial t^2} + (RC + GL) \frac{\partial e}{\partial t} + RGe \tag{8}$$

subject to the boundary conditions

$$e(0,t) = E_0 \cos \omega t \qquad e(x,t) \text{ bounded as } x \to \infty \tag{9}$$

If we assume a product solution $e(x,t) = X(x)T(t)$ and separate variables, we obtain

$$\frac{X''}{X} = \frac{LCT'' + (RC + GL)T' + RGT}{T} = \mu$$

Thus the factor T satisfies the equation

$$LCT'' + (RC + GL)T' + (RG - \mu)T = 0$$

and hence T must be of one or the other of the forms

$$e^{pt}(A \cos qt + B \sin qt)$$

$$e^{pt}(At + B)$$

$$Ae^{p_1 t} + Be^{p_1 t}$$

Under no circumstances can the last two expressions represent periodic behavior Moreover, the first expression can represent periodic behavior only if $p = 0$, which is impossible, since $p = -(RC + GL)/2LC \neq 0$. Hence, no product solution of Eq. (8) is capable of describing what we know the steady-state behavior of the line must be.

If we reconsider the problem, in an attempt to find an alternative method of solution, it seems reasonable to expect that, under the given conditions, the voltage along the line will vary harmonically with time while exhibiting attenuation and phase shift depending on the distance from the sending end. Hence we are led to try an expression of the form

(10) $$e(x,t) = E_0e^{-ax}\cos(\omega t + bx)$$

If $a > 0$, this obviously satisfies each of the boundary conditions (9), and perhaps the constants a and b can be determined so that it will satisfy the differential equation also.

If we substitute the tentative solution (10) into the telephone equation (8), divide out E_0e^{-ax} and collect terms, without difficulty we obtain

$$[a^2 - b^2 + LC\omega^2 - RG]\cos(\omega t + bx) + [2ab + \omega(RC + GL)]\sin(\omega t + bx) = 0$$

This will be an identity if and only if

(11) $$a^2 - b^2 = RG - LC\omega^2$$

(12) $$2ab = -(RC + GL)\omega$$

Now by adding the square of Eq. (12) to the square of Eq. (11), we obtain

$$(a^2 + b^2)^2 = (RG - LC\omega^2)^2 + (RC + GL)^2\omega^2$$

or

(13) $$a^2 + b^2 = \sqrt{(RG - LC\omega^2)^2 + (RC + GL)^2\omega^2}$$

Finally, by solving (11) and (13) simultaneously, we find

$$a^2 = \tfrac{1}{2}\left[\sqrt{(RG - LC\omega^2)^2 + (RC + GL)^2\omega^2} + (RG - LC\omega^2)\right]$$

$$b^2 = \tfrac{1}{2}\left[\sqrt{(RG - LC\omega^2)^2 + (RC + GL)^2\omega^2} - (RG - LC\omega^2)\right]$$

From the form of these equations it is clear that a^2 and b^2 are positive. Hence a and b are real, and, with their values now determined, Eq. (10) becomes the required solution. In a similar manner, of course, the steady-state response to a signal voltage of the form $E_0 \sin \omega t$ can be found.

By means of these results it is now possible to find the steady-state voltage corresponding to *any* periodic signal voltage. For if $e(0,t) = f(t)$ is a periodic function with period $2p$, then it can be expanded in a Fourier series,

$$f(t) = \frac{a_0}{2} + a_1\cos\frac{\pi t}{p} + a_2\cos\frac{2\pi t}{p} + \cdots + b_1\sin\frac{\pi t}{p} + b_2\sin\frac{2\pi t}{p} + \cdots$$

and the steady-state solution for each of these terms can be found. Then, since the telephone equation is linear, the sum of the steady-state responses to each of these terms will be the steady-state response of the line to the entire signal $f(t)$. Moreover, if the input signal is not periodic, the steady-state solution can still be found by a similar analysis using the Fourier integral rather than a Fourier series.

EXAMPLE 4

A very thin sheet of metal coincides with the square in the xy-plane whose vertices are the points (0,0), (1,0), (1,1), and (0,1). The upper and lower faces of the sheet are perfectly insulated, so that heat flow in it is purely two-dimensional. Initially, the temperature distribution in the sheet is $u(x,y,0) = f(x,y)$. If there are no sources of heat in the sheet, find the temperature at any point at any subsequent time, given that the edges parallel to the x-axis are perfectly insulated and that the edges parallel to the y-axis are maintained at the constant temperature 0°.

Here we have to solve the two-dimensional form of the heat equation [Eq. (14), Sec. 8.2],

(14) $$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = a^2\frac{\partial u}{\partial t}$$

subject to the boundary conditions

$$u(0,y,t) = 0 \qquad u(1,y,t) = 0 \tag{15}$$

$$\left.\frac{\partial u}{\partial y}\right|_{x,0,t} = 0 \qquad \left.\frac{\partial u}{\partial y}\right|_{x,1,t} = 0 \tag{16}$$

and the initial condition

$$u(x,y,0) = f(x,y) \tag{17}$$

Because we now have three independent variables, we begin with a product solution of the form

$$u(x,y,t) = X(x)Y(y)T(t) \tag{18}$$

Then substituting this into Eq (14) and attempting to separate variables, we get

$$\frac{X''}{X} = a^2\frac{T'}{T} - \frac{Y''}{Y} \tag{19}$$

Although y and t enter together on the right-hand side of (19), they are both independent of x, and so each side of the equation must be a constant, say μ. Thus the factor X satisfies the equation

$$X'' = \mu X$$

If $\mu > 0$, say $\mu = \lambda^2$, we have

$$X = A \cosh \lambda x + B \sinh \lambda x$$

But from the first of the boundary conditions (15), namely,

$$u(0,y,t) = X(0)Y(y)T(t) = AY(y)T(t) = 0$$

it follows that $A = 0$ Likewise, from the second of the conditions (15), namely,

$$u(1,y,t) = X(1)Y(y)T(t) = (B \sinh \lambda)Y(y)T(t) = 0$$

it follows that $B = 0$. Hence, when $\mu > 0$, the factor $X(x)$ vanishes identically, and only a trivial solution is possible

If $\mu = 0$, we have

$$X = Ax + B$$

and again the boundary conditions (15) can be satisfied only if $A = B = 0$

Finally, if $\mu < 0$, say $\mu = -\lambda^2$, we have

$$X = A \cos \lambda x + B \sin \lambda x$$

From the first of the boundary conditions (15) we conclude that $A = 0$. The second condition requires that $(B \sin \lambda)Y(y)T(t) = 0$ and, since we cannot permit B to be zero, we must have

$$\sin \lambda = 0 \qquad \text{and} \qquad \lambda = m\pi \qquad m = 1, 2, 3, \ldots$$

Therefore,

$$X_m(x) = \sin m\pi x \qquad m = 1, 2, 3, \ldots \tag{20}$$

Continuing with the other equation arising from (19), we now have

$$a^2\frac{T'}{T} - \frac{Y''}{Y} = -m^2\pi^2$$

or

$$\frac{Y''}{Y} = a^2\frac{T'}{T} + m^2\pi^2 \tag{21}$$

Since y and t are also independent, each member of the last equation must be a constant, say η Thus, the factor Y satisfies the equation

$$Y'' = \eta Y$$

If $\eta > 0$, say $\eta = \nu^2$, we have

$$Y = C \cosh \nu y + D \sinh \nu y$$

and $$Y' = \nu C \sinh \nu y + \nu D \cosh \nu y$$

But, from the first of the boundary conditions (16), namely,

$$\left.\frac{\partial u}{\partial y}\right|_{x,0,t} \equiv X(x)Y'(0)T(t) = X(x)(\nu D)T(t) = 0$$

it follows that $D = 0$. Likewise, from the second of the conditions (16), namely,

$$\left.\frac{\partial u}{\partial y}\right|_{x,1,t} \equiv X(x)Y'(1)T(t) = X(x)(\nu C \sinh \nu)T(t) = 0$$

it follows that $C = 0$ Hence, when $\eta > 0$, the factor $Y(y)$ vanishes identically, and only a trivial solution is possible

If $\eta = 0$, we have

$$Y = Cy + D$$

and this time the boundary conditions (16) require that $C = 0$ but do not restrict D Hence, $Y = D$ is a possible solution for the factor Y

Finally, if $\eta < 0$, say $\eta = -\nu^2$, we have

$$Y = C \cos \nu y + D \sin \nu y$$

and $$Y' = -\nu C \sin \nu y + \nu D \cos \nu y$$

From the first of the conditions (16) we conclude again that $D = 0$ The second of the conditions (16) requires that

$$X(x)(-\nu C \sin \nu)T(t) = 0$$

and, since we cannot permit $C = 0$, we must have

$$\sin \nu = 0 \qquad \text{and} \qquad \nu = n\pi \qquad n = 1, 2, 3,$$

Therefore, $$Y_n(y) = \cos n\pi y \qquad n = 1, 2, 3,$$

or, including the solution $Y = \text{constant}$ obtained when $\eta = 0$,

(22) $$Y_n(y) = \cos n\pi y \qquad n = 0, 1, 2, 3,$$

From (21) it is now clear that the factor T satisfies the equation

$$T' = -\frac{(m^2 + n^2)\pi^2}{a^2} T$$

and, hence, that

(23) $$T = E_{mn}e^{-[(m^2+n^2)\pi^2/a^2]t}$$

Therefore, combining (20) and (22) with (23) we can write the product solution (18) explicitly

(24) $$u_{mn}(x,y,t) = E_{mn} \sin m\pi x \cos n\pi y \; e^{[(m^2+n^2)\pi^2/a^2]t}$$

None of the product solutions (24), by itself, can reduce to the required initial temperature distribution (17) Hence, we must form a series of these and attempt to make this series satisfy the initial temperature condition But now, since we have two independent parameters m and n in the product solutions, the general solution for u will be a *double* series

$$u(x,y,t) = \sum_{m,n} u_{mn}(x,y,t) = \sum_{n=0}^{\infty} \sum_{m=1}^{\infty} E_{mn} \sin m\pi x \cos n\pi y \; e^{-[(m^2+n^2)\pi^2/a^2]t}$$

When $t = 0$, this must reduce to $f(x,y)$ that is,

(25) $$f(x,y) = \sum_{n=0}^{\infty} \left(\sum_{m=1}^{\infty} E_{mn} \sin m\pi x \right) \cos n\pi y$$

Now the inner summation in (25) is a function only of n and x, say $G_n(x)$, and, hence, (25) can be written

$$f(x,y) = \sum_{n=0}^{\infty} G_n(x) \cos n\pi y$$

But, for any particular value of x, this is just the Fourier half-range cosine expansion of $f(x,y)$, thought of now as a function of y for $0 \leqq y \leqq 1$ Hence, by familiar theory, we can write

$$(26) \qquad G_n(x) = 2 \int_0^1 f(x,y) \cos n\pi y \, dy$$

But, by definition,

$$G_n(x) = \sum_{m=1}^{\infty} E_{mn} \sin m\pi x$$

and this is just the half-range sine expansion of the now known function $G_n(x)$ for $0 \leqq x \leqq 1$ Hence,

$$(27) \qquad E_{mn} = 2 \int_0^1 G_n(x) \sin m\pi x \, dx$$

If we wish, we can substitute for $G_n(x)$ from (26) into (27), getting

$$\begin{aligned} E_{mn} &= 2 \int_0^1 \left[2 \int_0^1 f(x,y) \cos n\pi y \, dy \right] \sin m\pi x \, dx \\ &= 4 \int_0^1 \int_0^1 f(x,y) \cos n\pi y \sin m\pi x \, dy \, dx \end{aligned}$$

With E_{mn} determined for all values of m and n, the formal solution is now complete

EXERCISES

1 Can the procedure illustrated in Example 1 be modified to obtain a description of the motion of the string when the frequency of the impressed force is one of the natural frequencies of the string? How?

2 Can the procedure illustrated in Example 1 be modified to obtain a description of the motion of the string when the impressed force is of the form $f(x,t) = \phi(x)\theta(t)$, where **(a)** $\theta(t)$ is periodic? How? **(b)** $\theta(t)$ is not periodic? How?

3 Work Example 1 with

$$\phi(x) = \begin{cases} 0 & 0 < x < \dfrac{l}{4} \\ 1 & \dfrac{l}{4} < x < \dfrac{3l}{4} \\ 0 & \dfrac{3l}{4} < x < l \end{cases}$$

4 Work Example 1 with $\phi(x) = x(l - x)$

5 A uniform string of length l is acted upon by a distributed frictional force equal at each point to $-\dfrac{cw}{g}\dfrac{\partial y}{\partial t}$, where c is an arbitrary proportionality constant Discuss the subsequent motion of the string given that it begins with initial displacement $y(x,0) = g(x)$ and initial velocity $y(x,0) = h(x)$ In particular, show that certain frequencies in the "spectrum" of the string may be overdamped, and determine which ones Do the concepts of magnification ratio and phase shift apply to the vibration of a string with viscous damping? How?

6 A uniform string for which $a = 1$ and $l = \pi$ begins to move with initial displacement $y(x,0) = g(x)$ and initial velocity $y(x,0) = h(x)$ Determine the subsequent motion of the

string if it is acted upon by a distributed frictional force equal at each point to $-\frac{cw}{g}\frac{\partial y}{\partial t}$,

where **a** $c = 1$ **b** $c = 2$ **c** $c = 4$ **d** $c = 8$

7 A uniform shaft of length l with both ends free, vibrating torsionally, is acted upon by a periodic impressed torque equal at each point to $(g/\rho J)\phi(x) \sin \omega t$ If the initial displacement of the shaft is $\theta(x,0) = g(x)$ and if its initial angular velocity is $\dot{\theta}(x,0) = h(x)$, discuss the problem of determining its subsequent motion

8 Work Exercise 7 for a uniform shaft of length l fixed at $x = 0$ and free at $x = l$

9 Find the steady-state motion produced in a uniform beam of length l which is simply supported at each end given that the beam is acted upon by a distributed load whose magnitude per unit length is $x(l - x) \sin \omega t$

10 A uniform cantilever beam is built in at $x = 0$ and free at $x = l$ Find the steady-state motion produced in the beam by a distributed load whose magnitude per unit length is $x \sin \omega t$

11 Work Example 2, given that the left end of the rod is perfectly insulated

12 A slender rod of infinite length has its curved surface perfectly insulated Find the steady-state temperature distribution in the rod if the temperature at the finite end of the rod varies according to the law $u(0,t) = \sin \omega t$ Explain how this result can be used to determine the steady-state temperature distribution produced by an arbitrary periodic temperature condition at the finite end of the rod

13 A slender rod of length l has its curved surface perfectly insulated Its right end is maintained at the constant temperature $u(l,t) = 0$ At the left end the temperature varies according to the law $u(0,t) = \sin \omega t$ Determine the steady state temperature distribution in the rod Explain how this result can be used to determine the steady-state temperature distribution produced in the rod by an arbitrary periodic temperature condition at the left end [Hint Verify that λ can be chosen so that

$$u_1(x,t) = \sin \omega t \cos \frac{\lambda x}{l} \cosh \lambda \left(2 - \frac{x}{l}\right) - \cos \omega t \sin \frac{\lambda x}{l} \sinh \lambda \left(2 - \frac{x}{l}\right)$$

and

$$u_2(x,t) = \sin \omega t \cos \lambda \left(2 - \frac{x}{l}\right) \cosh \frac{\lambda x}{l} - \cos \omega t \sin \lambda \left(2 - \frac{x}{l}\right) \sinh \frac{\lambda x}{l}$$

are solutions of the one-dimensional heat equation Then determine A_1 and A_2 so that $u(x,t) = A_1u_1(x,t) + A_2u_2(x,t)$ satisfies the boundary conditions of the problem]

14 A slender rod of length l has its curved surface and left end perfectly insulated Heat is generated within the rod at a rate per unit volume equal to $\phi(x)$ Find the temperature in the rod as a function of x and t, if the right end of the rod is maintained at the constant temperature $u(l,t) = 0$ and if the initial temperature distribution in the rod is $u(x,0) = g(x)$

15 If the transmission line in Example 3 is initially "dead," i e , if at $t = 0$ the potential and current along the line are identically zero, determine the complete response of the line, transient as well as steady-state, to the signal voltage $E_0 \cos \omega t$ [Hint Show that, if $-p \pm iq$ are the roots of the equation

$$LCm^2 + (RC + GL)m + (RG + \lambda^2) = 0$$

then $$u_\lambda = e^{-pt}[A(\lambda) \cos qt + B(\lambda) \sin qt] \sin \lambda x$$

is a solution of the telephone equation which is bounded as $x \to \infty$ and is zero for all values of t when $x = 0$ Then show that the steady-state solution found in Example 3 plus the integral of u_λ over all values of λ is a solution which satisfies both boundary conditions (9) Finally, determine $A(\lambda)$ and $B(\lambda)$ so that both e and $\frac{\partial e}{\partial t}$ are zero when $t = 0$]

16 Work Example 3, by replacing the signal voltage $E_0 \cos \omega t$ by $E_0 e^{i\omega t}$ and assuming a solution of the form $u(x,t) = E_0 e^{i\omega t+(a+ib)x}$

17 Work Example 4, given that the edges from (0,0) to (0,1) and (1,0) are maintained at the constant temperature 0° and the other two edges are insulated

18 Determine E_{mn} in Exercise 4 for $f(x,y) = x + y$

19 A thin sheet of metal coincides with the square in the xy-plane whose vertices are the points (0,0), (1,0), (1,1), and (0,1) Along the edge from (0,0) to (1,0) the temperature distribution $u(x,0) = f(x)$ is maintained The other three edges are maintained at the temperature 0° Find the steady-state temperature as a function of x and y

20 Work Exercise 19 if the boundary conditions are

a $u(x,0) = f(x)$ $\qquad \left.\dfrac{\partial u}{\partial y}\right|_{x,1} = \left.\dfrac{\partial u}{\partial x}\right|_{0,y} = \left.\dfrac{\partial u}{\partial x}\right|_{1,y} = 0$

b $u(x,0) = u(x,1) = 0°$ $\qquad \left.\dfrac{\partial u}{\partial x}\right|_{0,y} = 0 \qquad u(1,y) = f(y)$

21 If an arbitrary temperature distribution exists along *each* of the edges of a square sheet of metal, how can the steady-state temperature distribution in the sheet be found?

22 A thin sheet of metal bounded by the x-axis, the lines $x = 0$ and $x = 1$, and stretching to infinity in the y-direction has its upper and lower faces insulated and its vertical edges maintained at the constant temperature 0° Over its base the temperature distribution $u(x,0) = 100°$ is maintained Find the steady-state temperature at any point in the sheet

23 Work Exercise 22 if the boundary conditions are

a $\left.\dfrac{\partial u}{\partial x}\right|_{0,y} = \left.\dfrac{\partial u}{\partial x}\right|_{1,y} = 0 \qquad u(x,0) = 100°$

b $u(0,y) = 0° \qquad u(1,y) = 100° \qquad u(x,0) = 100x$

c $u(0,y) = 0° \qquad \left.\dfrac{\partial u}{\partial x}\right|_{1,y} = 0 \qquad u(x,0) = 100°$

24 Work Exercise 22, given that the left edge and the lower edge of the sheet are maintained at the temperature 0° and the known distribution $u(1,y) = f(y)$ is maintained along the right edge

25 Determine the natural frequencies and nodal lines of a uniform square drumhead

8.7

Laplace transform methods

In Chap 7 we observed how the Laplace transformation converted an ordinary, linear, constant-coefficient differential equation into a linear algebraic equation from which the transform of the dependent variable could readily be found. In much the same way, the Laplace transformation can often be used to advantage in solving linear, constant-coefficient partial differential equations in two independent variables. In such cases it leads not to an *algebraic* equation but to an *ordinary differential* equation in the transform of the dependent variable The general procedure is as follows

The given partial differential equation, with its accompanying boundary conditions and initial conditions, is transformed with respect to one of its independent variables, usually t Partial derivatives with respect to this variable are, of course, transformed by the familiar formulas of Theorem 2 of Sec. 7 2 and its corollary. For partial derivatives with respect to the other independent variable we assume* that the operations of differentiating and taking the Laplace transform can be interchanged Then, if the independent variables are x and t, say, we have

$$\mathcal{L}\left\{\frac{\partial f(x,t)}{\partial x}\right\} = \int_0^\infty \frac{\partial f(x,t)}{\partial x} e^{-st}\,dt = \frac{\partial}{\partial x}\int_0^\infty f(x,t)e^{-st}\,dt = \frac{d}{dx}\mathcal{L}\{f(x,t)\}$$

the derivative in the last term being a total derivative because $\mathcal{L}\{f(x,t)\}$ is not a function of t. Similar formulas, of course, hold for x-derivatives of higher orders Thus, the result of the transformation is an ordinary differential equation in $\mathcal{L}\{f(x,t)\}$ in which x is the independent variable and s enters as a parameter. Because s occurs in the coefficients of the differential equation, the arbitrary constants appearing in its complete solution will in general be functions of s which must be determined by imposing the transformed boundary conditions on the complete solution of the transformed differential equation After this has been done, the inverse transformation is carried out and the solution to the original problem is obtained The details of this process can best be made clear through examples

EXAMPLE 1

A semi-infinite string is initially at rest in a position coinciding with the positive half of the x-axis At $t = 0$, the left end of the string begins to move along the y-axis in a manner described by $y(0,t) = f(t)$, where $f(t)$ is a known function Find the displacement $y(x,t)$ of the string at any point at any subsequent time

The partial differential equation to be solved is, of course, the one-dimensional wave equation

(1) $$\frac{\partial^2 y}{\partial t^2} = a^2\frac{\partial^2 y}{\partial x^2}$$

subject to the boundary conditions

(2) $$y(0,t) = f(t)$$

(3) $y(x,t)$ bounded as $x \to \infty$

and the initial conditions

(4) $$y(x,0) = 0$$

(5) $$\left.\frac{\partial y}{\partial t}\right|_{x,0} = 0$$

If we take the Laplace transform of Eq (1) *with respect to t*, we obtain

$$s^2\mathcal{L}\{y(x,t)\} - sy(x,0) - \left.\frac{\partial y}{\partial t}\right|_{x,0} = a^2\mathcal{L}\left\{\frac{\partial^2 y(x,t)}{\partial x^2}\right\} = a^2\frac{d^2}{dx^2}\mathcal{L}\{y(x,t)\}$$

or, using the initial conditions (4) and (5),

(6) $$\frac{d^2\mathcal{L}\{y(x,t)\}}{dx^2} - \frac{s^2}{a^2}\mathcal{L}\{y(x,t)\} = 0$$

---

* This is justified by Theorems 3 and 6, Sec 7.1.

Solving this ordinary differential equation for $\mathcal{L}\{y(x,t)\}$, we find without difficulty that

$$\mathcal{L}\{y(x,t)\} = A(s)e^{-(s/a)x} + B(s)e^{(s/a)x} \tag{7}$$

To determine the coefficient functions $A(s)$ and $B(s)$, we observe first that, if $y(x,t)$ remains finite as $x \to \infty$ [condition (3)], so must $\mathcal{L}\{y(x,t)\}$. Hence, $B(s)$ must be zero. Furthermore, putting $x = 0$ in (7) after $B(s)$ is set equal to zero, we have $\mathcal{L}\{y(0,t)\} = A(s)$, and, from the boundary condition (2), we have $\mathcal{L}\{y(0,t)\} = \mathcal{L}\{f(t)\}$. Therefore, (7) becomes

$$\mathcal{L}\{y(x,t)\} = \mathcal{L}\{f(t)\}e^{-(s/a)x}$$

The inverse of this can be found at once by suppressing the exponential factor and using Corollary 2 of Theorem 6, Sec. 7.4. The solution to our problem is, therefore,

$$y(x,t) = f\left(t - \frac{x}{a}\right)u\left(t - \frac{x}{a}\right)$$

which represents a wave traveling to the right along the string with velocity a. Evidently, the effect of this wave is to give the string at a general point the same displacement that the left end of the string had x/a units of time earlier.

EXAMPLE 2

A semi-infinite string is initially at rest in a position coinciding with the positive half of the x-axis. A concentrated transverse force of magnitude F_0 moves along the string with constant velocity v, beginning at $t = 0$ at the point $x = 0$. Find the displacement $y(x,t)$ of the string at any point at any subsequent time.

In this problem, since there is an external force applied to the string, we must use the nonhomogeneous wave equation [Eq. (2), Sec. 8.2]

$$\frac{\partial^2 y}{\partial t^2} = a^2 \frac{\partial^2 y}{\partial x^2} + \frac{g}{w} F(x,t)$$

To obtain $F(x,t)$ we observe that a single concentrated load F_0 acting at the point $x = vt$ corresponds to a load per unit length which is infinite at $x = vt$ and zero everywhere else. Hence, since F_0 is assumed to act on the string in the negative y-direction,

$$F(x,t) = -F_0\, \delta\left(t - \frac{x}{v}\right)$$

where $\delta(t - x/v)$ is the unit impulse, or δ function, which we discussed in Sec. 7.7. Our problem, therefore, is to solve the equation

$$\frac{\partial^2 y}{\partial t^2} = a^2 \frac{\partial^2 y}{\partial x^2} - \frac{g}{w} F_0\, \delta\left(t - \frac{x}{v}\right) \tag{8}$$

subject to the boundary conditions

$$y(0,t) = 0 \tag{9}$$

$$y(x,t) \text{ bounded as } x \to \infty \tag{10}$$

and the initial conditions

$$y(x,0) = 0 \tag{11}$$

$$\left.\frac{\partial y}{\partial t}\right|_{x,0} = 0 \tag{12}$$

If we take the Laplace transform of Eq. (8) with respect to t and use the initial conditions (11) and (12), we obtain, just as in Example 1,

$$s^2\mathcal{L}\{y(x,t)\} = a^2 \frac{d^2}{dx^2}\mathcal{L}\{y(x,t)\} - \frac{g}{w} F_0\, e^{-(s/v)x}$$

or

$$\frac{d^2}{dx^2}\mathcal{L}\{y(x,t)\} - \frac{s^2}{a^2}\mathcal{L}\{y(x,t)\} = \frac{gF_0}{a^2 w} e^{-(s/v)x} \tag{13}$$

The solution of this equation by the methods of Chap 2 presents no difficulty, and we find, for the complete solution,

$$\mathcal{L}\{y(x,t)\} = A(s)e^{-(s/a)x} + B(s)e^{(s/a)x} + \begin{cases} \dfrac{gv^2F_0}{w(a^2 - v^2)s^2} e^{-(s/v)x} & v \neq a \\ -\dfrac{gF_0}{2was} xe^{-(s/v)x} & v = a \end{cases} \tag{14}$$

In each case we must have $B(s) = 0$ in order that $\mathcal{L}\{y(x,t)\}$ should remain finite as $x \to \infty$. To determine $A(s)$ we have, from the boundary condition (9), the information that, when $x = 0$,

$$\mathcal{L}\{y(x,t)\} = \mathcal{L}\{y(0,t)\} = 0$$

Hence, substituting into Eq (14), we obtain

$$A(s) = \begin{cases} -\dfrac{gv^2F_0}{w(a^2 - v^2)s^2} & v \neq a \\ 0 & v = a \end{cases}$$

and, therefore,

$$\mathcal{L}\{y(x,t)\} = \begin{cases} \dfrac{gv^2F_0}{w(a^2 - v^2)s^2}[e^{-(x/v)s} - e^{-(x/a)s}] & v \neq a \\ -\dfrac{gF_0}{2was} xe^{-(x/a)s} & v = a \end{cases}$$

Taking inverses, we have finally

$$y(x,t) = \frac{gv^2F_0}{w(a^2 - v^2)}\left[\left(t - \frac{x}{v}\right)u\left(t - \frac{x}{v}\right) - \left(t - \frac{x}{a}\right)u\left(t - \frac{x}{a}\right)\right] \qquad v \neq a \tag{15}$$

and

$$y(x,t) = -\frac{gF_0}{2wa} x u\left(t - \frac{x}{a}\right) \qquad v = a \tag{16}$$

Plots of (15) in the "subsonic" case $v = \frac{3}{4}a$, and the "supersonic" case $v = \frac{5}{4}a$, and of the "transonic" case $v = a$ described by (16) are shown in Fig 8 14 for a typical time t The dis-

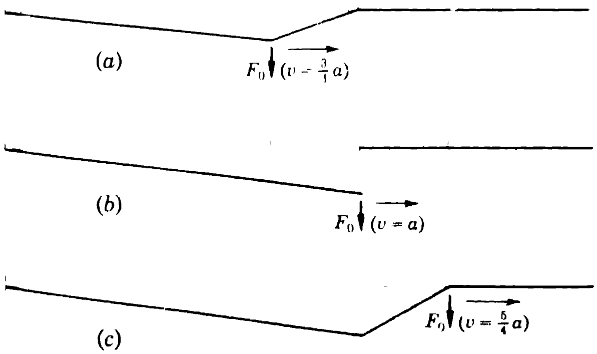


FIGURE 8 14 Plot showing the displacement of a semi-infinite string produced by a concentrated force moving along the string with a velocity of (a) $\frac{3}{4}$, (b) 1, and (c) $\frac{5}{4}$ times the propagation velocity for the string continuity in $y(x,t)$ in the "transonic" case when the disturbance travels with exactly the propagation velocity a is interesting.

EXAMPLE 3

A semi-infinite cable of negligible leakage and inductance is initially "dead." At $t = 0$ an arbitrary signal voltage $E(t)$ is suddenly applied at the sending end. Find the potential $e(x,t)$ at any point on the cable at any subsequent time.

In this problem we have to solve the telegraph equation (21*a*), Sec. 8.2,

$$(17) \qquad \frac{\partial^2 e}{\partial x^2} = a^2 \frac{\partial e}{\partial t}\dagger \qquad a^2 = RC$$

subject to the boundary conditions

$$(18) \qquad e(0,t) = E(t)$$

$$(19) \qquad e(x,t) \text{ bounded as } x \to \infty$$

and the initial condition

$$(20) \qquad e(x,0) = 0$$

Taking the Laplace transform of (17) with respect to t and using the initial condition (20), we obtain

$$\frac{d^2}{dx^2}\mathcal{L}\{e(x,t)\} = a^2 s\mathcal{L}\{e(x,t)\}$$

as the ordinary differential equation satisfied by the transform of the potential. Solving this for $\mathcal{L}\{e(x,t)\}$ we find without difficulty that

$$(21) \qquad \mathcal{L}\{e(x,t)\} = A(s)e^{-a\sqrt{s}\,x} + B(s)e^{a\sqrt{s}\,x}$$

Since $e(x,t)$ and, hence, $\mathcal{L}\{e(x,t)\}$ are to remain finite as $x \to \infty$, it is necessary that $B(s) = 0$. To determine $A(s)$ we observe that, when $x = 0$,

$$\mathcal{L}\{e(x,t)\} = \mathcal{L}\{E(t)\}$$

Hence, substituting into Eq. (21), we find

$$A(s) = \mathcal{L}\{E(t)\}$$

and

$$(22) \qquad \mathcal{L}\{e(x,t)\} = \mathcal{L}\{E(t)\}e^{-ax\sqrt{s}}$$

To determine $e(x,t)$ it will be necessary to use the convolution theorem, but, before this can be done, we must know the inverse of $e^{-ax\sqrt{s}}$. Up to this point in our work we have not encountered any function of t having this function of s for its transform. However, it can be shown (see Exercises 1 and 2) that

$$\mathcal{L}\left\{\frac{be^{-b^2/4t}}{2\sqrt{\pi}\,t^{3/2}}\right\} = e^{-b\sqrt{s}}$$

Hence, taking $b = ax$ and setting up the convolution integral, we obtain from (22)

$$(23) \qquad e(x,t) = \frac{ax}{2\sqrt{\pi}}\int_0^t E(t-\lambda)\frac{e^{-a^2x^2/4\lambda}}{\lambda\sqrt{\lambda}}\,d\lambda$$

In particular, if $E(t)$ is a unit step voltage, we have, since $u(t-\lambda) = 1$ for $\lambda < t$, and

---

† This is identical with the one-dimensional heat equation, and so all our conclusions apply equally well to the problem of the flow of heat in a slender, insulated, semi-infinite rod whose left end is maintained at the time-dependent temperature $u_0(t)$.

$u(t - \lambda) = 0$ for $\lambda > t$,

$$e(x,t) = \frac{ax}{2\sqrt{\pi}} \int_0^t \frac{e^{-a^2x^2/4\lambda}}{\lambda\sqrt{\lambda}}\, d\lambda$$

If we let $a^2x^2/4\lambda = z^2$, then $\lambda = a^2x^2/4z^2$, $d\lambda = -a^2x^2/2z^3\, dz$, and the last integral becomes

$$e(x,t) = \frac{ax}{2\sqrt{\pi}} \int_\infty^{ax/2\sqrt{t}} e^{-z^2} \frac{8z^3}{a^3x^3}\left(-\frac{a^2x^2}{2z^3}\, dz\right)$$

$$= \frac{2}{\sqrt{\pi}} \int_{ax/2\sqrt{t}}^\infty e^{-z^2}\, dz$$

(24) $$= \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-z^2}\, dz - \frac{2}{\sqrt{\pi}} \int_0^{ax/2\sqrt{t}} e^{-z^2}\, dz$$

Under the substitution $z^2 = v$, the first integral becomes

$$\frac{1}{\sqrt{\pi}} \int_0^\infty e^{-v} v^{1/2-1}\, dv = \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right) = 1$$

since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Hence, Eq. (24) can be written

$$e(x,t) = 1 - \frac{2}{\sqrt{\pi}} \int_0^{ax/2\sqrt{t}} e^{-z^2}\, dz$$

(25) $$= 1 - \operatorname{erf} \frac{ax}{2\sqrt{t}}$$

where

(26) $$\operatorname{erf}(\theta) = \frac{2}{\sqrt{\pi}} \int_0^\theta e^{-z^2}\, dz$$

This is the so-called **error function**, a tabulated function which can be found in most handbooks of mathematical tables.*

EXERCISES

1. If

$$f(\lambda) = \int_0^\infty \frac{e^{-z}e^{-\lambda/z}}{\sqrt{z}}\, dz$$

show by means of the substitution $u = \lambda/z$ that

$$f(\lambda) = \sqrt{\lambda} \int_0^\infty \frac{e^{-u}e^{-\lambda/u}}{u^{3/2}}\, du$$

Hence, by differentiating the first expression for $f(\lambda)$, show that

$$f'(\lambda) = -\frac{f(\lambda)}{\sqrt{\lambda}}$$

---

* Actually, most handbooks list not the error function as here defined and used in physics and engineering, but rather the so-called **probability integral** of mathematical statistics,

$$\Phi(\theta) = \frac{1}{\sqrt{2\pi}} \int_0^\theta e^{-w^2/2}\, dw$$

If the substitution $z = w/\sqrt{2}$ is made in the error function (26), it becomes

$$\frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{2}\,\theta} e^{-w^2/2}\, dw$$

and we obtain the relation

$$\operatorname{erf} \theta = 2\Phi(\sqrt{2}\,\theta)$$

Solve this differential equation, using the fact that

$$f(0) = \Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

and show that

$$f(\lambda) = \sqrt{\pi}\, e^{-2\sqrt{\lambda}}$$

Finally, use this result to show that

$$\mathcal{L}\left\{\frac{e^{-b^2/4t}}{\sqrt{\pi t}}\right\} = \frac{e^{-b\sqrt{s}}}{\sqrt{s}}$$

2 Use the results of the last exercise, together with Theorem 8, Sec. 7.4, to show that

$$\mathcal{L}\left\{\frac{be^{-b^2/4t}}{2\sqrt{\pi}\, t^{3/2}}\right\} = e^{-b\sqrt{s}}$$

3 a In Example 3, what is the response of the line if $E(t)$ is a unit impulse voltage? (Hint. Recall from Sec. 7.7 the relation between the response of a system to a unit step function and to a unit impulse.)

b Using Eq. (25) and the appropriate Duhamel formula, obtain a formula different from Eq. (23) for the response of the line in Example 3 to a general voltage.

4 Using Laplace transform methods, determine the motion of a string of length l whose initial displacement and initial velocity are, respectively, $y(x,0) = \sin m\pi x/l$ and $\dot{y}(x,0) = \sin n\pi x/l$. Can these results be used to obtain the motion of the string produced by arbitrary initial conditions? How?

5 Using Laplace transform methods, determine the response of a string of length l to a distributed force $f(x,t) = (\sin n\pi x/l) \sin \omega t$ if the string is initially at rest in its equilibrium position. Explain how these results can be used to determine the response of the string to a distributed force $f(x,t) = g(x) \sin \omega t$, where $g(x)$ is defined arbitrarily on the interval $0 < x < l$, and to a distributed force $(\sin n\pi x/l)\, h(t)$, where $h(t)$ is an arbitrary periodic function whose frequency is distinct from each natural frequency of the string.

6 Work Example 2, given that the transverse force which moves along the string is a rectangular pulse of height F_0 initially acting on the portion of the string between $x = 0$ and $x = 1$.

7 A semi-infinite string whose weight per unit length is w has its left end fixed at the origin. The infinite end is fastened to a ring which slides without friction along a vertical rod. Initially, the string is at rest in a position coinciding with the positive x-axis. At $t = 0$ the support which maintained the string in its horizontal position is removed and the string begins to fall freely under the influence of gravity. Determine its subsequent position as a function of x and t.

8 A shaft of uniform cross section is built in at $x = 0$ and free at $x = l$. At $t = 0$, while the shaft is at rest in its equilibrium position, a constant torque T_0 is suddenly applied to the free end. Find the Laplace transform of the resultant angular displacement. What is the angular displacement of the free end as a function of time? (Hint. The boundary condition at $x = l$ is $E_sJ \dfrac{\partial \theta}{\partial x} = T_0$.)

9 Work Exercise 8, given that the torque applied at the free end is a unit impulse instead of a step function.

10 A semi-infinite string initially at rest in a position coinciding with the positive x-axis is acted upon by a concentrated force $F_0 \sin \omega t$ applied at the point $x = b$. Find the Laplace transform of the resultant displacement of the string. What is the displacement of the string at the point $x = b$ as a function of time?

CHAPTER NINE

Bessel Functions and Legendre Polynomials

9.1

Theoretical preliminaries

In solving partial differential equations by the method of separation of variables, we are often led to ordinary differential equations with variable coefficients which cannot be solved in terms of familiar functions. The usual procedure in such cases is to obtain solutions in the form of infinite series, which can be taken as the definitions of new functions to be studied in detail and eventually tabulated if they prove of sufficient importance. In this section we shall discuss the general problem of obtaining series solutions of the form

$$y = (x - a)^r[a_0 + a_1(x - a) + a_2(x - a)^2 + \cdots] \tag{1}$$

for the general linear second-order differential equation

$$y'' + P(x)y' + Q(x)y = 0 \tag{2}$$

We shall not require the exponent r to be a positive integer, and in general it will not be. Hence the solutions we obtain will usually not be Taylor expansions.

The analysis involves a consideration of several cases, depending upon the behavior of the coefficient functions $P(x)$ and $Q(x)$ at the point $x = a$ around which we propose to expand the solution y In most of our work, the variables x and y and the coefficient functions $P(x)$ and $Q(x)$ will all be real. However, this is not a necessary restriction, and, in the basic definitions and theorems we shall introduce in this section, x, y, $P(x)$, and $Q(x)$ may be either real or complex. In the first place, both $P(x)$ and $Q(x)$ may be analytic at $x = a$; that is, they may possess Taylor expansions around the point $x = a$. When this happens, $x = a$ is said to be an **ordinary point** of the differential equation. A point which is not an ordinary point is called a **singular point**. At a singular point, although $P(x)$ and $Q(x)$ do not both possess

Taylor expansions, it may be that the products

$$(x-a)P(x) \qquad \text{and} \qquad (x-a)^2Q(x)$$

do have Taylor expansions. A singular point at which this is the case is said to be **regular**; otherwise it is called **irregular.** In our work we shall be concerned exclusively with the expansion of solutions of Eq (2) around ordinary points and regular singular points

EXAMPLE 1

For the differential equation

$$y'' + \frac{2}{x}y' + \frac{3}{x(x-1)^3}y = 0$$

$x = 0$ and $x = 1$ are singular points, since at $x = 0$ both $P(x)$ and $Q(x)$ become infinite, while at $x = 1$, although $P(x)$ is analytic, $Q(x)$ becomes infinite All other points are ordinary points The point $x = 0$ is a regular singular point, since each of the products

$$xP(x) = x\frac{2}{x} = 2 \qquad \text{and} \qquad x^2Q(x) = x^2\frac{3}{x(x-1)^3} = \frac{3x}{(x-1)^3}$$

is analytic at $x = 0$, i e , can be expanded in a series of positive integral powers of x The point $x = 1$ is an irregular singular point, however, because, although the product

$$(x-1)P(x) = (x-1)\left(\frac{2}{x}\right) = \frac{2(x-1)}{x}$$

is analytic at $x = 1$, the product

$$(x-1)^2Q(x) = (x-1)^2\frac{3}{x(x-1)^3} = \frac{3}{x(x-1)}$$

becomes infinite there and hence is not analytic

The importance of the classification of values of x into ordinary and singular points is apparent from the following theorems, which are proved in more advanced treatments of the theory of differential equations.*

THEOREM 1

At an ordinary point $x = a$ of the differential equation $y'' + P(x)y' + Q(x)y = 0$, every solution is analytic; i.e , can be represented by a series of the form

$$y = a_0 + a_1(x-a) + a_2(x-a)^2 +$$

Moreover, the radius of convergence of each series solution is equal to the distance from a to the nearest singular point of the equation.

THEOREM 2

At a regular singular point $x = a$ of the differential equation

$$y'' + P(x)y' + Q(x)y = 0$$

---

* See, for instance, E. T Whittaker and G. N. Watson, "Modern Analysis," pp. 194–203, The Macmillan Company, New York, 1943.

there is at least one solution which possesses an expansion of the form

$$y = (x - a)^r[a_0 + a_1(x - a) + a_2(x - a)^2 + \cdots]$$

and this series will converge for $0 < |x - a| < R$, where R is the distance from a to the nearest of the other singular points of the equation

THEOREM 3

At an irregular singular point $x = a$ of the differential equation

$$y'' + P(x)y' + Q(x)y = 0$$

there are in general no solutions with expansions consisting solely of powers of $x - a$.

In using Theorems 1 and 2 to infer the radius of convergence of power series solutions of Eq. (2), it must be borne in mind that the singular point nearest to, but distinct from, the point of expansion may be complex, even though the point around which we are expanding is real. For instance, for the differential equation

$$y'' + \frac{1}{1 + x^2}y' + y = 0$$

the coefficient functions

$$P(x) = \frac{1}{1 + x^2} \qquad \text{and} \qquad Q(x) = 1$$

are analytic for all real values of x. However, $P(x)$ fails to be analytic at $x = \pm i$; hence, these two points are singular points of the differential equation. Therefore, a series solution around the ordinary point $x = 2$, say, would have radius of convergence $R = \sqrt{5}$, since in the complex plane the distance from the point of expansion $x = 2$ to the nearest singular point $x = i$ (or $x = -i$) is $\sqrt{5}$ (Fig. 9.1).

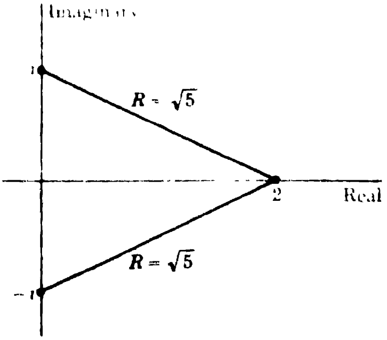


FIGURE 9.1 Plot showing the radius of convergence as the distance to a complex singular point

To obtain series solutions of Eq. (2) around an ordinary point or a reqular singular point we use the so-called **method of Frobenius.*** First of all, for convenience, we translate axes, if necessary, so that the point of expansion becomes the point $x = 0$. Now, if $x = 0$ is either an ordinary point or a regular singular point, both $xP(x)$ and $x^2Q(x)$ are analytic, and, hence, we can write

---

* Named for the German mathematician F. G. Frobenius (1849–1917).

$$xP(x) = b_0 + b_1x + b_2x^2 + \cdots$$
$$x^2Q(x) = c_0 + c_1x + c_2x^2 + \cdots$$

Therefore, multiplying Eq. (2) by x^2 and then substituting for $xP(x)$ and $x^2Q(x)$, we have

(3) $$x^2y'' + x(b_0 + b_1x + b_2x^2 + \cdots)y' + (c_0 + c_1x + c_2x^2 + \cdots)y = 0$$

Next, we assume a series of the desired form

(4) $$y = x^r(a_0 + a_1x + a_2x^2 + \cdots)$$

where, without loss of generality, we can suppose that $a_0 \neq 0$. If we substitute this series into Eq. (3), we have

$$\begin{aligned} x^2[a_0r(r-1)x^{r-2} + a_1(r+1)rx^{r-1} + a_2(r+2)(r+1)x^r + \cdots] \\ + x(b_0 + b_1x + b_2x^2 + \cdots)[a_0rx^{r-1} + a_1(r+1)x^r \\ + a_2(r+2)x^{r+1} + \cdots] + (c_0 + c_1x + c_2x^2 + \cdots) \\ \times (a_0x^r + a_1x^{r+1} + a_2x^{r+2} + \cdots) = 0 \end{aligned}$$

or, collecting terms on the various powers of x,

(5) $$\begin{aligned} a_0[r(r-1) + b_0r + c_0]x^r + \{a_1[(r+1)r + b_0(r+1) + c_0] \\ + a_0(b_1r + c_1)\}x^{r+1} + \{a_2[(r+2)(r+1) + b_0(r+2) + c_0] \\ + a_1[b_1(r+1) + c_1] + a_0(b_2r + c_2)\}x^{r+2} + \cdots = 0 \end{aligned}$$

Equation (5) will be an identity if and only if the coefficient of each power of x is zero, and thus we obtain the set of equations

(6) $$\begin{aligned} a_0[r(r-1) + b_0r + c_0] &= 0 \\ a_1[(r+1)r + b_0(r+1) + c_0] + a_0(b_1r + c_1) &= 0 \\ a_2[(r+2)(r+1) + b_0(r+2) + c_0] + a_1[b_1(r+1) + c_1] + a_0(b_2r + c_2) &= 0 \\ \cdots\cdots\cdots\cdots &\cdot \end{aligned}$$

Since $a_0 \neq 0$, it follows from the first of these equations that

(7) $$r^2 + (b_0 - 1)r + c_0 = 0$$

This quadratic equation in r is known as the **indicial equation** of the differential equation relative to the point of expansion, and its roots r_1 and r_2 are known as the **exponents** of the differential equation at that point. For each of these values there is, in general, a series solution of the form (4). And the coefficients in these expansions can be determined, one by one, from the successive equations in the set (6), which express each of the a's, in turn, in terms of the a's preceding it in the series (4).

EXAMPLE 2

Find series solutions for the equation $9x^2y'' + (x + 2)y = 0$ around the origin

Since $P(x) = 0$ and $Q(x) = (x + 2)/9x^2$, it follows that the origin is a regular singular point of the given equation Hence, by Theorem 2, there exists at least one solution with an expansion of the form

$$y = x^r(a_0 + a_1x + a_2x^2 + \cdots)$$

Substituting this into the differential equation, we have

$$\begin{aligned}
&9x^2[a_0r(r-1)x^{r-2} + a_1(r+1)rx^{r-1} + \cdots + a_{k+1}(r+k+1)(r+k)x^{r+k-1} + \cdots]\\
&+ x(\qquad a_0x^r + \cdots + a_kx^{r+k} + \cdots)\\
&+ 2(\quad a_0x^r + a_1x^{r+1} + \cdots + a_{k+1}x^{r+k+1} + \cdots) = 0
\end{aligned}$$

or, collecting terms,

$$\begin{aligned}
a_0[9r(r-1) + 2]x^r + \{a_1[9(r+1)r + 2] + a_0\}x^{r+1} + \cdots\\
+ \{a_{k+1}[9(r+k+1)(r+k) + 2] + a_k\}x^{r+k+1} + \cdots = 0
\end{aligned}$$

For this to be an identity we must have

$$\begin{aligned}
9r(r-1) + 2 &= 0\\
a_1[9(r+1)r + 2] + a_0 &= 0\\
&\cdots\\
a_{k+1}[9(r+k+1)(r+k) + 2] + a_k &= 0\\
&\cdot
\end{aligned}$$

The first of these is the indicial equation whose roots are $r = \frac{1}{3}, \frac{2}{3}$ From the second we find that

$$a_1 = -\frac{a_0}{(3r+1)(3r+2)}$$

and, from the general recurrence relation, we have

$$a_{k+1} = -\frac{a_k}{[3(r+k)+1][3(r+k)+2]}$$

Considering these first for $r = \frac{1}{3}$ and then for $r = \frac{2}{3}$, we obtain the coefficient sequences

$$r = \frac{1}{3}: \quad a_0 = a_0, \quad a_1 = -\frac{a_0}{2\cdot 3}, \quad a_2 = -\frac{a_1}{5\cdot 6} = \frac{a_0}{2\cdot 3\cdot 5\cdot 6},$$

$$a_3 = -\frac{a_2}{8\cdot 9} = -\frac{a_0}{2\cdot 3\cdot 5\cdot 6\cdot 8\cdot 9},$$

$$r = \frac{2}{3} \quad a_0 = a_0, \quad a_1 = -\frac{a_0}{3\cdot 4}, \quad a_2 = -\frac{a_1}{6\cdot 7} = \frac{a_0}{3\cdot 4\cdot 6\cdot 7},$$

$$a_3 = -\frac{a_2}{9\cdot 10} = -\frac{a_0}{3\cdot 4\cdot 6\cdot 7\cdot 9\cdot 10},$$

With these coefficients, taking $a_0 = 1$ for convenience, we can construct the two particular solutions

$$y_1 = x^{1/3}\left(1 - \frac{x}{2\cdot 3} + \frac{x^2}{2\cdot 3\cdot 5\cdot 6} - \frac{x^3}{2\cdot 3\cdot 5\cdot 6\cdot 8\cdot 9} + \cdots\right)$$

$$y_2 = x^{2/3}\left(1 - \frac{x}{3\cdot 4} + \frac{x^2}{3\cdot 4\cdot 6\cdot 7} - \frac{x^3}{3\cdot 4\cdot 6\cdot 7\cdot 9\cdot 10} + \cdots\right)$$

Since all values of x except $x = 0$ are ordinary points of the given differential equation, it follows from Theorem 2 that these series converge for all values of x Finally, since y_1 and y_2 are clearly independent, i e , have nonvanishing Wronskian, it follows from Theorem 2, Sec. 2 1, that the complete solution of the given equation is an arbitrary linear combination of these two particular solutions

If the indicial equation has a double root, it is obvious that two series solutions cannot be obtained by the present method It is also true (though not obvious) that, if the roots of the indicial equation differ by an integer, this method fails, in general, to provide a second series solution.* In either of these cases, however, a second solution can be found by the method of Sec 2 1, that is, by assuming $y = \phi(x)y_1(x)$, where $y_1(x)$ is the first series solution, and then determining $\phi(x)$ so that the product will satisfy the given differential equation

EXERCISES

1 Find the singular points of each of the following equations, and determine whether they are regular or irregular.

a $xy'' + y' + y = 0$

b $x^2y'' + y' + y = 0$

c $x^2(1 - x)y'' + (1 - x)y' + y = 0$

d $(1 - x^2)y'' + y' + y = 0$

2 Find the indicial equation relative to each of the singular points of each of the equations in Exercise 1

Find series solutions around the origin for each of the following equations

3 $4x^2y'' + (4x + 1)y = 0$

4 $x^2y'' + (x - 2)y = 0$

5 $2x^2y'' + 3xy' + (x^2 - 1)y = 0$

6 $2x^2y'' + (2x^2 + 3x)y' + (x - 1)y = 0$

7 The "point at infinity" is said to be an ordinary point or a singular point of the differential equation

$$\frac{d^2y}{dx^2} + P(x)\frac{dy}{dx} + Q(x)y = 0$$

according as the equation obtained from this by the substitution $x = 1/u$ has an ordinary point or a singular point at $u = 0$ Show that, under this transformation, the original equation becomes

$$u^4\frac{d^2y}{du^2} + \left[2u^3 - u^2P\left(\frac{1}{u}\right)\right]\frac{dy}{du} + Q\left(\frac{1}{u}\right)y = 0$$

and use this result to determine the nature of the point at infinity for the equation

$$(x^2 + 1)y'' + y' + y = 0$$

8 Show that, if the origin is an irregular singular point of the differential equation (2), then the indicial equation relative to the origin is at most of the first degree

9 Verify that, if the roots of the indicial equation differ by unity, then, in general, the two roots lead to the same series solution of Eq (2). When will this not be the case? Is this true if the roots differ by an integer greater than 1?

10 Verify that under the change of dependent variable defined by the substitution

$$y = ze^{-\frac{1}{2}\int P(x)dx}$$

the differential equation (2) becomes

$$\frac{d^2z}{dx^2} + R(x)z = 0$$

where $R(x) = Q(x) - \frac{1}{2}\frac{dP(x)}{dx} - \frac{1}{4}P^2(x)$

---

* See Exercise 9.

9.2

The series solution of Bessel's equation

Probably the most important of all variable-coefficient differential equations is

(1)
$$x^2\frac{d^2y}{dx^2}+x\frac{dy}{dx}+(\lambda^2x^2-\nu^2)y=0$$

which is known as **Bessel's equation of order ν with parameter λ**.† This arises in a great variety of problems, including almost all applications involving partial differential equations, such as the wave equation or the heat equation, in regions having circular symmetry

As a preliminary step in the solution of Eq (1), let us change the independent variable from x to t by means of the substitution

(2)
$$t=\lambda x$$

Since $\quad \dfrac{dy}{dx}=\lambda\dfrac{dy}{dt} \quad$ and $\quad \dfrac{d^2y}{dx^2}=\lambda^2\dfrac{d^2y}{dt^2}$

Eq (1) then becomes

(3)
$$t^2\frac{d^2y}{dt^2}+t\frac{dy}{dt}+(t^2-\nu^2)y=0$$

which is known simply as **Bessel's equation of order ν.**

For Eq (3) it is clear that

$$P(t)=\frac{1}{t} \qquad \text{and} \qquad Q(t)=\frac{t^2-\nu^2}{t^2}$$

Hence, the origin is a regular singular point of the equation, and all other values of t are ordinary points At the origin, where we propose to obtain series solutions of (3), the indicial equation [Eq. (7), Sec 9 1] is $r^2-\nu^2=0$, and, therefore, by the theory of the preceding section, we are led to try a series solution of the form

(4)
$$y=t^\nu(a_0+a_1t+a_2t^2+\qquad)$$

Substituting this series into Eq (3) and displaying the terms in a convenient array, we have

$$\begin{array}{llllll}
 & [a_0\nu(\nu-1)t^\nu & +\,a_1(\nu+1)\nu t^{\nu+1} & +\,a_2(\nu+2)(\nu+1)t^{\nu+2}+ & +\,a_k(\nu+k)(\nu+k-1)t^{\nu+k}+ &] \\
+ & [\quad a_0\nu t^\nu & +\,a_1(\nu+1)t^{\nu+1} & +\,a_2(\nu+2)t^{\nu+2}+ & +\,a_k(\nu+k)t^{\nu+k}+ &] \\
+ & (& & a_0t^{\nu+2}+ & +\,a_{k-2}t^{\nu+k}+ &) \\
+ & (\;-\nu^2a_0t^\nu & -\,\nu^2a_1t^{\nu+1} & -\,\nu^2a_2t^{\nu+2}- & -\,\nu^2a_kt^{\nu+k}+ &)=0
\end{array}$$

This will be an identity if and only if the coefficient of every power of t is zero The coefficient of t^ν is automatically zero, since ν is a root of the indicial equation From the coefficient of $t^{\nu+1}$

---

† Named for the German mathematician and astronomer Friedrich Wilhelm Bessel (1784–1846), although special cases of this equation had been studied earlier by Jakob Bernoulli (1703), Daniel Bernoulli (1732), and Leonhard Euler (1764)

we obtain the condition

$$a_1(2\nu + 1) = 0$$

and, in general, for $k \geqq 2$, we obtain from the coefficient of $t^{\nu+k}$

$$a_k[(\nu + k)(\nu + k - 1) + (\nu + k) - \nu^2] + a_{k-2} \equiv a_k k(2\nu + k) + a_{k-2} = 0$$

or

$$a_k = -\frac{a_{k-2}}{k(2\nu + k)} \tag{5}$$

Now it is clear that a_1 must be zero for all values of ν except possibly $\nu = -\frac{1}{2}$, and even in this case we can assume $a_1 = 0$, since we are interested only in conditions *sufficient* for the existence of solutions of the form

$$\sum_{k=0}^{\infty} a_k t^{\nu+k}$$

Moreover, from (5) it is apparent that any coefficient a_k is a multiple of the second preceding coefficient a_{k-2} Hence, beginning with a_1, it follows that every coefficient with an odd subscript must vanish

On the other hand, starting with a_0, which is still perfectly arbitrary, and taking $k = 2, 4, 6,$ successively in the recurrence formula (5), we have

$$a_0 = a_0$$

$$a_2 = -\frac{a_0}{2(2\nu + 2)} = -\frac{a_0}{2^2 \cdot 1!(\nu + 1)}$$

$$a_4 = -\frac{a_2}{4(2\nu + 4)} = -\frac{a_2}{2^2 \cdot 2(\nu + 2)} = \frac{a_0}{2^4 \cdot 2!(\nu + 2)(\nu + 1)}$$

$$a_6 = -\frac{a_4}{6(2\nu + 6)} = -\frac{a_4}{2^2 \cdot 3(\nu + 3)} = -\frac{a_0}{2^6 \cdot 3!(\nu + 3)(\nu + 2)(\nu + 1)}$$

and, in general, $$a_{2m} = \frac{(-1)^m a_0}{2^{2m} m!(\nu + m) \cdots (\nu + 2)(\nu + 1)}$$

Now a_{2m} is the coefficient of $t^\nu t^{2m} \equiv t^{\nu+2m}$ in the series (4) for y. Hence it would be convenient if a_{2m} contained the factor $2^{\nu+2m}$ in its denominator instead of just 2^{2m} To achieve this, we write

$$a_{2m} = \frac{(-1)^m}{2^{\nu+2m} m!(\nu + m) \cdots (\nu + 2)(\nu + 1)} (2^\nu a_0)$$

Furthermore, the factors

$$(\nu + m) \cdots (\nu + 2)(\nu + 1)$$

suggest a factorial. In fact, if ν were an integer, a factorial could be created by multiplying numerator and denominator by $\nu!$ However, since ν is not necessarily an integer, we must use not $\nu!$ but its generalization $\Gamma(\nu + 1)$ (Sec. 7.3) for this purpose.

Then, except for the values

$\nu = -1, -2, -3, \ldots$

for which $\Gamma(\nu+1)$ is not defined, we can write

$$a_{2m} = \frac{(-1)^m}{2^{\nu+2m}m!(\nu+m)\cdots(\nu+2)(\nu+1)\Gamma(\nu+1)}[2^\nu\Gamma(\nu+1)a_0]$$

Since the gamma function satisfies the recurrence relation

$$z\Gamma(z) = \Gamma(z+1)$$

the expression for a_{2m} becomes finally

$$a_{2m} = \frac{(-1)^m}{2^{\nu+2m}m!\Gamma(\nu+m+1)}[2^\nu\Gamma(\nu+1)a_0]$$

Since a_0 is arbitrary and since we are looking only for particular solutions, we choose

$$a_0 = \frac{1}{2^\nu\Gamma(\nu+1)}$$

so that $$a_{2m} = \frac{(-1)^m}{2^{\nu+2m}m!\Gamma(\nu+m+1)}$$

The series for y is, therefore, from (4),

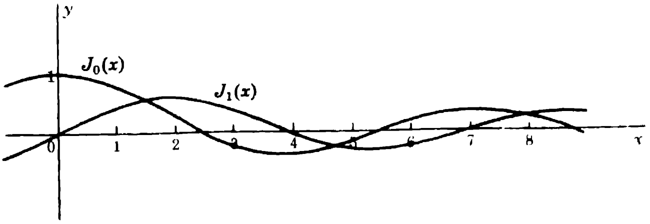
$$y(t) = t^\nu\left[\frac{1}{2^\nu\Gamma(\nu+1)} - \frac{t^2}{2^{\nu+2}\Gamma(\nu+2)} + \frac{t^4}{2^{\nu+4}2!\Gamma(\nu+3)} - \cdots\right]$$

$$(6) \qquad = \sum_{m=0}^{\infty}\frac{(-1)^m t^{\nu+2m}}{2^{\nu+2m}m!\Gamma(\nu+m+1)}$$

The function defined by this infinite series is known as the **Bessel function of the first kind of order ν** and is denoted by the symbol $J_\nu(t)$. Since Bessel's equation of order ν has no finite singular points except the origin, it follows from Theorem 2, Sec. 9.1, that the series for $J_\nu(t)$ converges for all values of t if $\nu \geq 0$. The graphs of $J_0(t)$ and $J_1(t)$ are shown in Fig. 9.2. Their resemblance to the graphs of $\cos t$ and $\sin t$ is interesting. In particular, they illustrate the important fact that for every value of ν the equation $J_\nu(t) = 0$ has infinitely many real roots.

Let us now consider the series arising from the other root of the indicial equation, namely, $r = -\nu$. We could, of course, begin again with a series analogous to (4) and determine its coefficients one by one, just as we did for $J_\nu(t)$. There is no need

FIGURE 9.2
Plot showing the Bessel functions of the first kind $J_0(x)$ and $J_1(x)$



to do this, however, for the final result can be obtained at once simply by replacing ν by $-\nu$ in the series (6), provided that the gamma functions appearing in the denominators of the various terms are all defined. This is necessarily the case unless ν is an integer, hence when ν is not an integer the function

(7) $$J_{-\nu}(t) = \sum_{m=0}^{\infty} \frac{(-1)^m t^{-\nu+2m}}{2^{-\nu+2m} m!\Gamma(-\nu+m+1)}$$

is a second particular solution of Bessel's equation of order ν. Moreover, since $J_{-\nu}(t)$ contains negative powers of t while $J_\nu(t)$ does not, it is obvious that in the neighborhood of the origin $J_{-\nu}(t)$ is unbounded while $J_\nu(t)$ remains finite. Hence $J_\nu(t)$ and $J_{-\nu}(t)$ cannot be proportional and, therefore, are two independent solutions of the Bessel equation According to Theorem 2, Sec 2.1, a complete solution of Bessel's equation when ν is not an integer is then

(8) $$y(t) = c_1 J_\nu(t) + c_2 J_{-\nu}(t)$$

Instead of $J_{-\nu}(t)$, some writers take the linear combination

(9) $$Y_\nu(t) = \frac{\cos \nu\pi J_\nu(t) - J_{-\nu}(t)}{\sin \nu\pi}$$

as a second, independent solution of Bessel's equation Using $Y_\nu(t)$, which is known as the **Bessel function of the second kind of order** ν, a complete solution of Bessel's equation can be written,

(10) $$y(t) = c_1 J_\nu(t) + c_2 Y_\nu(t) \qquad \nu \text{ not an integer}$$

In some applications it is convenient to use still another form of the general solution of Bessel's equation This is based upon the two particular solutions

(11) $$\begin{aligned} H_\nu^{(1)}(t) &= J_\nu(t) + iY_\nu(t) \\ H_\nu^{(2)}(t) &= J_\nu(t) - iY_\nu(t) \end{aligned}$$

These are known as **Hankel functions*** or **Bessel functions of the third kind of order** ν, and in terms of them a complete solution of Eq (3) can be written,

(12) $$y(t) = c_1 H_\nu^{(1)}(t) + c_2 H_\nu^{(2)}(t) \qquad \nu \text{ not an integer}$$

It is interesting to note that (8), (10), and (12) are correct expressions for the general solution of Eq (3) even when ν is an odd multiple of $\frac{1}{2}$ and the roots of the indicial equation $r^2 - \nu^2 = 0$ differ by an integer. In the last section we pointed out that, when this happens, a second, independent series solution of the form (4) will usually not exist It *may* exist, however, and Bessel's equation is one of the instances when it actually does †

---

* Named for the German mathematician Hermann Hankel (1839–1873).

† See Exercise 9, Sec. 9 1.

If ν is an integer, say $\nu = n$, the situation is somewhat different. Again the roots of the indicial equation differ by an integer, namely, $2n$, and it is to be expected that a second solution of the form (4) will not exist. In fact, considering $J_{-n}(t)$ as the limit of $J_\nu(t)$ as ν approaches $-n$ and remembering that when its argument approaches any nonpositive integer the gamma function becomes infinite, it follows that as ν approaches $-n$, the first n terms in the series (6) approach zero and the series effectively begins with the term for which $m = n$.

$$J_{-n}(t) = \sum_{m=n}^{\infty} \frac{(-1)^m t^{-n+2m}}{2^{-n+2m} m!\,\Gamma(-n+m+1)}$$

In this, let the variable of summation be changed from m to j by the substitution $m = j + n$. Then

$$\begin{aligned} J_{-n}(t) &= \sum_{j=0}^{\infty} \frac{(-1)^{j+n} t^{-n+2(j+n)}}{2^{-n+2(j+n)}(j+n)!\,\Gamma[-n+(j+n)+1]} \\ &= \sum_{j=0}^{\infty} \frac{(-1)^n(-1)^j t^{n+2j}}{2^{n+2j}\Gamma(n+j+1)j!} \\ &= (-1)^n J_n(t) \end{aligned}$$

Thus, when ν is an integer n, the function $J_{-\nu}(t)$ is proportional to $J_\nu(t)$. These two solutions are therefore not independent, and the linear combination $c_1J_\nu(t) + c_2J_{-\nu}(t)$ is no longer a complete solution of Bessel's equation. Moreover, without additional definitions, neither (10) nor (12) provides a complete solution, since $Y_\nu(t)$, as defined by (9), assumes the indeterminate form $0/0$ when ν is an integer.

A complete solution when ν is an integer can be found in either of several ways. One is to use the method developed in Sec. 2.1 for finding a second solution of a linear second-order differential equation when one solution is known. The result, as given by Eq. (5) of Sec. 2.1 with $y_1(t) = J_n(t)$ and $P(t) = 1/t$, is

$$y(t) = cJ_n(t) \int \frac{dt}{tJ_n^2(t)} + kJ_n(t)$$

The usual procedure, however, is to obtain a second, independent solution by evaluating the limit of $Y_\nu(t)$ as $\nu \to n$. The details are somewhat involved, and we shall not present them here. The limit function, which exists and is independent of $J_n(t)$ for all values of n, is commonly denoted by $Y_n(t)$, that is,

(13) $$Y_n(t) = \lim_{\nu \to n} Y_\nu(t) = \lim_{\nu \to n} \frac{\cos \nu\pi J_\nu(t) - J_{-\nu}(t)}{\sin \nu\pi}$$

The corresponding specializations of the Hankel functions (11) are defined in the obvious way in terms of $Y_n(t)$:

(14) $$\begin{aligned} H_n^{(1)}(t) &= J_n(t) + iY_n(t) \\ H_n^{(2)}(t) &= J_n(t) - iY_n(t) \end{aligned}$$

With Formulas (13) and (14), we can now eliminate from (10) and (12) the restriction that ν is not an integer and use these results for all values of ν, integral as well as nonintegral. Plots of $Y_0(t)$ and $Y_1(t)$ are shown in Fig. 9.3. Among other things, they illustrate the important fact that, for all values of ν, $Y_\nu(t)$ is unbounded in the neighborhood of the origin and has infinitely many real zeros.

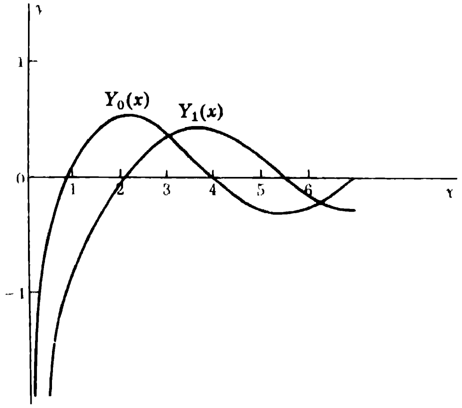


FIGURE 9.3 Plot showing the Bessel functions of the second kind $Y_0(x)$ and $Y_1(x)$.

Reversing the transformation (2) which we used to eliminate the parameter λ from the general form of Bessel's equation (1), we can now summarize the results of the preceding discussion in the following theorem

THEOREM 1

For all values of ν, a complete solution of Bessel's equation of order ν with parameter λ,

$$x^2y'' + xy' + (\lambda^2x^2 - \nu^2)y = 0$$

can be written in either of the forms

$$y(x) = c_1J_\nu(\lambda x) + c_2Y_\nu(\lambda x)$$
$$y(x) = c_1H_\nu^{(1)}(\lambda x) + c_2H_\nu^{(2)}(\lambda x)$$

If ν is not an integer, a complete solution can also be written

$$y(x) = c_1J_\nu(\lambda x) + c_2J_{-\nu}(\lambda x)$$

$J_\nu(\lambda x)$, $J_{-\nu}(\lambda x)$, and $Y_\nu(\lambda x)$ all have infinitely many real zeros If $\nu \geqq 0$, $J_\nu(\lambda x)$ is finite for all values of x, but $J_{-\nu}(\lambda x)$ and $Y_\nu(\lambda x)$ are unbounded in the neighborhood of the origin. $H_\nu^{(1)}(\lambda x)$ and $H_\nu^{(2)}(\lambda x)$ are complex-valued functions when x is real.

EXERCISES

1 If y_1 and y_2 are any two solutions of Bessel's equation of order ν, show that $y_1y_2' - y_1'y_2 = c/x$, where c is a suitable constant (Hint: Recall Abel's identity from Sec 2.1.)

2 By determining the coefficient of $1/x$ on the left-hand side, show that

$$J_\nu(x)J'_{-\nu}(x) - J'_\nu(x)J_{-\nu}(x) = -\frac{2}{\pi x}\sin \nu\pi$$

[Hint Use the result of Exercise 1 and the fact that $\Gamma(z)\Gamma(1-z) = \pi/(\sin \pi z)$ if z is not an integer]

3 If ν is not an integer, show that $J_\nu(x)$ and $J_{-\nu}(x)$ have no zeros in common (Hint Use the result of Exercise 2)

4 Show that

$$Y = \frac{\pi}{2\sin\nu\pi}\left[J_\nu(x)\int f(s)J_{-\nu}(s)\,ds - J_{-\nu}(x)\int f(s)J_\nu(s)\,ds\right]$$

is a particular integral of the nonhomogeneous Bessel equation

$$x^2y'' + xy' + (x^2 - \nu^2)y = xf(x)$$

if ν is not an integer (Hint Use the method of variation of parameters and the results of Exercise 2)

5 Show that, under the transformation $y = u/\sqrt{t}$, Bessel's equation of order ν becomes

$$\frac{d^2u}{dt^2} + \left(1 - \frac{4\nu^2 - 1}{4t^2}\right)u = 0$$

Hence show that, for large values of t, solutions of Bessel's equation are approximately described by expressions of the form

$$c_1\frac{\sin t}{\sqrt{t}} + c_2\frac{\cos t}{\sqrt{t}}$$

[More precisely, it can be shown that

$$J_\nu(t) \sim \sqrt{\frac{2}{\pi t}}\cos\left(t - \frac{\pi}{4} - \frac{\nu\pi}{2}\right)$$

$$Y_\nu(t) \sim \sqrt{\frac{2}{\pi t}}\sin\left(t - \frac{\pi}{4} - \frac{\nu\pi}{2}\right)$$

where the symbol $\sim$ means that the limit of the ratio of the two quantities connected by it approaches 1 as t becomes infinite]

9.3

Modified Bessel functions

There are certain equations closely resembling Bessel's equation which occur so often that their solutions are also named and studied as functions in their own right. The most important of these is

(1) $$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} - \left(1 + \frac{\nu^2}{x^2}\right)y = 0$$

which is known as the **modified Bessel equation of order ν**. Since this can be written in the form

$$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} + \left(i^2 - \frac{\nu^2}{x^2}\right)y = 0$$

it is evident that this is nothing but Bessel's equation of order ν with the imaginary parameter $\lambda = i$. However, in actual applications, to write the complete solution of (1) in the form

$$y = c_1 J_\nu(ix) + c_2 Y_\nu(ix)$$

and retain the imaginaries would be about as awkward as to take the solution of

$$\frac{d^2y}{dx^2} - y = 0$$

to be $\quad y = c_1 \cos ix + c_2 \sin ix$

and use this complex expression instead of resorting to real exponentials or hyperbolic functions. Accordingly, we seek modifications of $J_\nu(ix)$ and $Y_\nu(ix)$ which will be real functions of real variables

Now, $$J_\nu(ix) = \sum_{k=0}^{\infty} \frac{(-1)^k (ix)^{\nu+2k}}{2^{\nu+2k} k! \Gamma(\nu + k + 1)}$$

$$= i^\nu \sum_{k=0}^{\infty} \frac{x^{\nu+2k}}{2^{\nu+2k} k! \Gamma(\nu + k + 1)}$$

Moreover, $J_\nu(ix)$ multiplied by any constant will also be a solution of the equation we are considering Hence, in particular, we can multiply it by $i^{-\nu}$, getting

$$i^{-\nu} J_\nu(ix) = \sum_{k=0}^{\infty} \frac{x^{\nu+2k}}{2^{\nu+2k} k! \Gamma(\nu + k + 1)}$$

This is a completely real function, identical with $J_\nu(x)$ except that its terms, instead of alternating in sign, are all positive This new function, which is related to $J_\nu(x)$ in the same way that cosh x and sinh x are related to cos x and sin x, is known as the **modified Bessel function of the first kind of order** ν, $I_\nu(x)$. If ν is not an integer, the function $I_{-\nu}(x)$ obtained from $I_\nu(x)$ by replacing ν by $-\nu$ throughout is a second, independent solution of Eq (1), whose complete solution can therefore be written

$$y = c_1 I_\nu(x) + c_2 I_{-\nu}(x)$$

On the other hand, instead of using $I_{-\nu}(x)$, many writers take the second solution of the modified Bessel equation to be the linear combination

$$K_\nu(x) = \frac{\pi}{2} \cdot \frac{I_{-\nu}(x) - I_\nu(x)}{\sin \nu\pi}$$

which is known as the **modified Bessel function of the second kind of order** ν If ν is not an integer, this is a well-defined solution which is clearly independent of $I_\nu(x)$ If ν is an integer n, this assumes the indeterminate form 0/0, but a tedious evaluation by L'Hospital's rule leads to a limiting expression

$$K_n(x) = \lim_{\nu \to n} K_\nu(x) = \lim_{\nu \to n} \frac{\pi}{2} \cdot \frac{I_{-\nu}(x) - I_\nu(x)}{\sin \nu\pi}$$

which is a solution independent of $I_n(x)$. This is a useful result, because, as we might expect, $I_\nu(x)$ and $I_{-\nu}(x)$ are not independent when ν is an integer. In fact, when $\nu = n$, we have the identity

$$(-1)^n J_{-n}(ix) = J_n(ix)$$

and then, by obvious steps,

$$\begin{aligned} (i^2)^n J_{-n}(ix) &= J_n(ix) \\ i^n J_{-n}(ix) &= i^{-n} J_n(ix) \\ I_{-n}(x) &= I_n(x) \end{aligned}$$

Plots of $I_0(x)$ and $I_1(x)$ are shown in Fig. 9.4, plots of $K_0(x)$ and $K_1(x)$ in Fig. 9.5. As these graphs illustrate, the modified Bessel functions have no real zeros except possibly $x = 0$. They also illustrate that, for $\nu \geqq 0$, $I_\nu(x)$ is finite at the origin, but $K_\nu(x)$, like $I_{-\nu}(x)$, becomes infinite as x approaches zero.

Just as the ordinary Bessel equation, so the modified Bessel equation frequently occurs in a form containing a parameter λ:

(2) $$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} - \left(\lambda^2 + \frac{\nu^2}{x^2}\right)y = 0$$

A complete solution of this is, of course,

$$y = c_1 I_\nu(\lambda x) + c_2 K_\nu(\lambda x) \qquad \nu \text{ unrestricted}$$

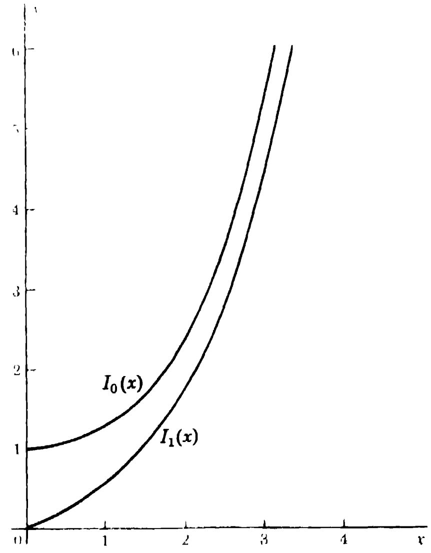


FIGURE 9.4 Plot showing the modified Bessel functions of the first kind $I_0(x)$ and $I_1(x)$.

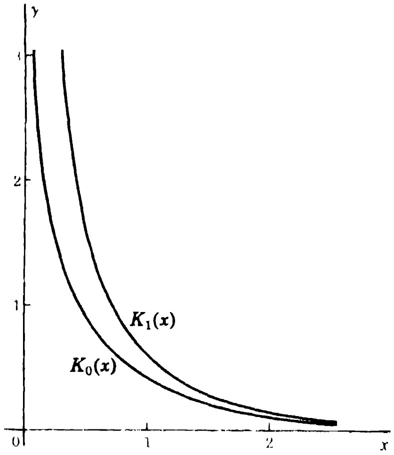


FIGURE 9.5 Plot showing the modified Bessel functions of the second kind $K_0(x)$ and $K_1(x)$.

If ν is not an integer, we have the alternative form

$$y = c_1 I_\nu(\lambda x) + c_2 I_{-\nu}(\lambda x)$$

A second equation closely related to Bessel's equation is

$$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} + \left(-i - \frac{\nu^2}{x^2}\right)y = 0 \tag{3}$$

This can be regarded either as Bessel's equation of order ν with parameter $\lambda = \pm\sqrt{-i}$ or as the modified Bessel equation of order ν with parameter $\lambda = \pm\sqrt{i}$ From the former point of view a complete solution can be written

$$y = c_1 J_\nu(\pm\sqrt{-i}\,x) + c_2 Y_\nu(\pm\sqrt{-i}\,x)$$

From the second point of view the solution can be written

$$y = d_1 I_\nu(\pm\sqrt{i}\,x) + d_2 K_\nu(\pm\sqrt{i}\,x)$$

Now a complete solution can be constructed from *any* pair of independent particular solutions, and it is customary in studying Eq (3) to select $J_\nu(\pm\sqrt{-i}\,x)$ and $K_\nu(\pm\sqrt{i}\,x)$ for this purpose The solution becomes unambiguous when a choice is made between the two square roots in each case Naturally enough, the positive, or principal, square roots are chosen Then since*

$$i = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2} = e^{i\pi/2} \qquad \text{and} \qquad -i = \cos\frac{3\pi}{2} + i\sin\frac{3\pi}{2} = e^{3i\pi/2}$$

it follows that

$$\sqrt{-i} = (e^{3i\pi/2})^{1/2} = (e^{i\pi/2})^{3/2} = i^{3/2}$$

---

* See Formula (7), Sec 14 7.

and we have for the complete solution

$$y = cJ_\nu(i^{3/2}x) + dK_\nu(i^{3/2}x)$$

Now

$$J_\nu(i^{3/2}x) = \sum_{k=0}^{\infty} \frac{(-1)^k (i^{3/2}x)^{\nu+2k}}{2^{\nu+2k}k!\Gamma(\nu+k+1)}$$

$$= i^{3\nu/2} \sum_{k=0}^{\infty} \frac{(-1)^k i^{3k} x^{\nu+2k}}{2^{\nu+2k}k!\Gamma(\nu+k+1)}$$

Moreover, i^{3k} can take on only one of the four values

$1 \qquad k = 0, 4, 8, \ldots$

$-i \qquad k = 1, 5, 9, \ldots$

$-1 \qquad k = 2, 6, 10, \ldots$

$i \qquad k = 3, 7, 11, \ldots$

Hence, the first, third, fifth, terms in $J_\nu(i^{3/2}x)$ are real, and the second, fourth, sixth, are imaginary. Separating the series into its real and imaginary parts, we obtain

$$J_\nu(i^{3/2}x) = i^{3\nu/2}\left[\sum_{j=0}^{\infty} \frac{(-1)^j x^{\nu+4j}}{2^{\nu+4j}(2j)!\Gamma(\nu+2j+1)} + i\sum_{j=0}^{\infty} \frac{(-1)^j x^{\nu+2+4j}}{2^{\nu+2+4j}(2j+1)!\Gamma(\nu+2j+2)}\right]$$

$$= i^{3\nu/2}\left[\sum\nolimits_r + i\sum\nolimits_i\right]$$

Furthermore,

$$i^{3\nu/2} = (e^{i\pi/2})^{3\nu/2} = e^{3i\pi\nu/4} = \cos\frac{3\pi\nu}{4} + i\sin\frac{3\pi\nu}{4}$$

and, therefore,

$$J_\nu(i^{3/2}x) = \left(\cos\frac{3\nu\pi}{4} + i\sin\frac{3\nu\pi}{4}\right)\left(\sum\nolimits_r + i\sum\nolimits_i\right)$$

$$= \left(\cos\frac{3\nu\pi}{4}\sum\nolimits_r - \sin\frac{3\nu\pi}{4}\sum\nolimits_i\right) + i\left(\cos\frac{3\nu\pi}{4}\sum\nolimits_i + \sin\frac{3\nu\pi}{4}\sum\nolimits_r\right)$$

$J_\nu(i^{3/2}x)$ thus consists of one purely real series plus i times a second purely real series. The series forming the real part of this expression defines the function **ber**$_\nu\, x$. The series forming the imaginary part defines the function **bei**$_\nu\, x$. The letters *be* suggest the relation between these new functions and the Bessel functions themselves. The terminal letters r and i, of course, suggest the adjectives *real* and *imaginary*. For the important case $\nu = 0$, we have explicitly

$$\text{ber}_0\,x \equiv \text{ber}\,x = \sum_{j=0}^{\infty} \frac{(-1)^j x^{4j}}{2^{4j}[(2j)!]^2}$$

$$\text{bei}_0\,x \equiv \text{bei}\,x = \sum_{j=0}^{\infty} \frac{(-1)^j x^{4j+2}}{2^{4j+2}[(2j+1)!]^2}$$

Plots of ber x and bei x are shown in Fig 9.6 The graphs oscillate with ever-increasing amplitudes.

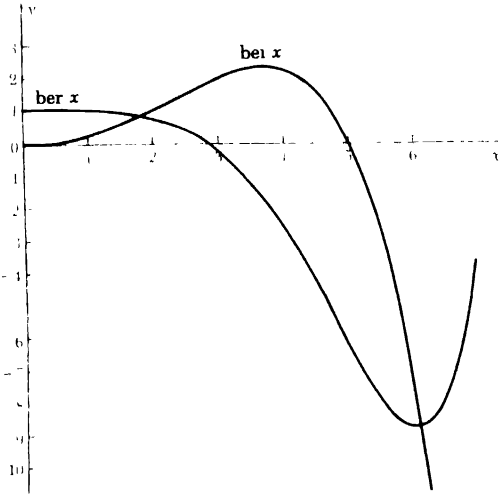


FIGURE 9.6 Plot showing the functions ber x and bei x

In a similar way the function $K_\nu(i^{1/2}x)$ can be expressed as a real series plus i times another real series These series are taken as the definitions of the new functions **ker**$_\nu$ x and **kei**$_\nu$ x, respectively. A complete solution of Eq (3) can thus be written

$$y = c(\text{ber}_\nu\,x + i\,\text{bei}_\nu\,x) + d(\text{ker}_\nu\,x + i\,\text{kei}_\nu\,x)$$

The function $\text{ber}_\nu\,x + i\,\text{bei}_\nu\,x$ is finite at the origin, but becomes infinite as x becomes infinite, $\text{ker}_\nu\,x + i\,\text{kei}_\nu\,x$ is infinite at the origin, but approaches zero as x becomes infinite

EXERCISES

1 Show that $J_\nu(i^{-3/2}x) = \text{ber}_\nu\,x - i\,\text{bei}_\nu\,x$

2 Show that $J_0(i^{1/2}x) = \text{ber}\,x - i\,\text{bei}\,x$ Is $J_\nu(i^{1/2}x) = \text{ber}_\nu\,x - i\,\text{bei}_\nu\,x$ in general?

3 Show that $(x\,\text{ber}'\,x)' = -x\,\text{bei}\,x$ and that $(x\,\text{bei}'\,x)' = x\,\text{ber}\,x$

4 Write out $\text{ber}_1\,x$ and $\text{bei}_1\,x$

5 Show that, under the transformation $y = u/\sqrt{x}$, the modified Bessel equation of order ν becomes

$$\frac{d^2u}{dx^2} - \left(1 + \frac{4\nu^2 - 1}{4x^2}\right)u = 0$$

Hence show that, for large values of x, solutions of the modified Bessel equation are approximately described by expressions of the form

$$c_1 \frac{e^{-x}}{\sqrt{x}} + c_2 \frac{e^{x}}{\sqrt{x}}$$

(More precisely, it can be shown that, as x becomes infinite,

$$I_\nu(x) \sim \frac{e^x}{\sqrt{2\pi x}} \qquad \text{and} \qquad K_\nu(x) \sim \sqrt{\frac{\pi}{2x}}\, e^{-x}\Big)$$

9.4

Equations reducible to Bessel's equation

There are many differential equations whose solutions can be expressed in terms of Bessel functions. In particular, we have the large and important family described in the following theorem.

THEOREM 1

If $(1 - a)^2 \geqq 4c$ and if neither d, p, nor q is zero, then, except in the obvious special cases when it reduces to Euler's equation,* the differential equation

$$x^2y'' + x(a + 2bx^p)y' + [c + dx^{2q} + b(a + p - 1)x^p + b^2x^{2p}]y = 0$$

has as a complete solution

$$y = x^\alpha e^{-\beta x^p}[c_1J_\nu(\lambda x^q) + c_2Y_\nu(\lambda x^q)]$$

where

$$\alpha = \frac{1-a}{2} \qquad \beta = \frac{b}{p} \qquad \lambda = \frac{\sqrt{|d|}}{q} \qquad \nu = \frac{\sqrt{(1-a)^2 - 4c}}{2q}$$

If $d < 0$, J_ν and Y_ν are to be replaced by I_ν and K_ν, respectively. If ν is not an integer, Y_ν and K_ν can be replaced by $J_{-\nu}$ and $I_{-\nu}$ if desired.

The proof of this theorem, while straightforward, is lengthy and involved, and we shall not present it here. It consists in transforming the given equation by means of the substitutions

$$y = x^{(1-a)/2}e^{-(b/p)x^p}Y \qquad \text{and} \qquad x = \left(\frac{qX}{\sqrt{|d|}}\right)^{1/q}$$

and verifying that, when the parameters are properly identified, the result is precisely Bessel's equation.

One special case of Theorem 1 is of sufficient interest to be stated as a corollary.

COROLLARY 1

If $(1 - r)^2 \geqq 4b$, if $a \neq 0$, and if either $s > r - 2$ or $b = 0$, then a complete solution of the equation

$$(x^ry')' + (ax^s + bx^{r-2})y = 0$$

is

$$y = x^\alpha[c_1J_\nu(\lambda x^\gamma) + c_2Y_\nu(\lambda x^\gamma)]$$

---

* Equation (10), Sec. 2.6.

where

$$\alpha = \frac{1-r}{2} \qquad \gamma = \frac{2-r+s}{2} \qquad \lambda = \frac{2\sqrt{|a|}}{2-r+s} \qquad \nu = \frac{\sqrt{(1-r)^2-4b}}{2-r+s}$$

If $a < 0$, J_ν and Y_ν are to be replaced by I_ν and K_ν, respectively If ν is not an integer, Y_ν and K_ν can be replaced by $J_{-\nu}$ and $I_{-\nu}$ if desired.

EXAMPLE 1

Find a complete solution of the equation

$$x^2y'' + x(4x^4 - 3)y' + (4x^8 - 5x^2 + 3)y = 0$$

Clearly, this is a special case of the equation of Theorem 1 with

$$a = -3 \qquad b = 2 \qquad p = 4 \qquad c = 3 \qquad d = -5 \qquad q = 1$$

Hence, $\alpha = 2 \qquad \beta = \frac{1}{2} \qquad \lambda = \sqrt{|-5|} = \sqrt{5} \qquad$ and $\qquad \nu = 1$

A complete solution is, therefore,

$$y = x^2e^{-x^4/2}[c_1I_1(\sqrt{5}\,x) + c_2K_1(\sqrt{5}\,x)]$$

EXAMPLE 2

What is a complete solution of the equation $y'' + y = 0$?

Obviously, one possibility is

$$y = c_1 \cos x + c_2 \sin x$$

However, $y'' + y = 0$ is also a special case of the equation of Corollary 1, with $r = 0$, $s = 0$, $a = 1$, and $b = 0$ Hence,

$$\alpha = \frac{1}{2} \qquad \gamma = 1 \qquad \lambda = 1 \qquad \nu = \frac{1}{2}$$

and so we can also write

$$y = d_1\sqrt{x}\,J_{1/2}(x) + d_2\sqrt{x}\,J_{-1/2}(x)$$

It follows, therefore, from Theorem 2, Sec 2 1, that, for proper choice of the constants c_1 and c_2, each of the particular solutions

$$\sqrt{x}\,J_{1/2}(x) \qquad \text{and} \qquad \sqrt{x}\,J_{-1/2}(x)$$

must be expressible in the form $c_1 \cos x + c_2 \sin x$

Now, since $\Gamma(\frac{3}{2}) = \frac{1}{2}\Gamma(\frac{1}{2}) = \frac{1}{2}\sqrt{\pi}$, the series for $J_{1/2}(x)$ begins with the term

$$\frac{x^{1/2}}{2^{1/2}\Gamma(\frac{3}{2})} = \sqrt{\frac{2x}{\pi}}$$

Hence, the series for $\sqrt{x}\,J_{1/2}(x)$ begins with the term $\sqrt{2/\pi}\,x$ Therefore, if we write

$$\sqrt{x}\,J_{1/2}(x) = \sqrt{\frac{2}{\pi}}\,x - \cdots = c_1 \cos x + c_2 \sin x$$

and put $x = 0$ in this identity, we find $c_1 = 0$ Subsequently, by equating the coefficients of x, we find

$$\sqrt{\frac{2}{\pi}} = c_2$$

We have thus established the interesting and important result that

$$\sqrt{x}\,J_{1/2}(x) = \sqrt{\frac{2}{\pi}}\sin x \qquad \text{or} \qquad J_{1/2}(x) = \sqrt{\frac{2}{\pi x}}\sin x$$

In a similar manner it can be shown that

$$J_{-\frac{1}{2}}(x) = \sqrt{\frac{2}{\pi x}} \cos x$$

EXERCISES

Find a complete solution of each of the following equations

1 $y'' + x^m y = 0$ 2 $xy'' + 2y' + 4xy = 0$

3 $x^2y'' + 3xy' + (1 + x)y = 0$ 4 $xy'' - y' + 4x^5y = 0$

5 $x^2y'' + 2x^2y' + (x^4 + x^2 - 2)y = 0$ 6 $x^2y'' + (2x^2 + x)y' + (x^2 + 3x - \ )y = 0$

7 Show that

$$I_{\frac{1}{2}}(x) = \sqrt{\frac{2}{\pi x}} \sinh x \qquad \text{and} \qquad I_{-\frac{1}{2}}(x) = \sqrt{\frac{2}{\pi x}} \cosh x$$

8 Show that any solution of

$$(x^{m-1}y')' = kx^{m-2}y \qquad \text{or} \qquad (x^{m-1}y')' = -kx^{m-2}y$$

will also satisfy the equation $(x^m y'')'' = k^2x^{m-2}y$

9 What is a complete solution of $(x^2y'')'' = 9y$?

10 What is a complete solution of $x^2y^{IV} + 8xy''' + 12y'' - y = 0$?

9.5

Identities for the Bessel functions

The Bessel functions are related by an amazing array of identities. Fundamental among these are the consequences of the following pair of theorems

THEOREM 1

$$\frac{d[x^\nu J_\nu(x)]}{dx} = x^\nu J_{\nu-1}(x)$$

PROOF To prove this theorem, we take the series for $J_\nu(x)$, multiply it by x^ν, and differentiate it term by term

$$J_\nu(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{\nu+2k}}{2^{\nu+2k}k!\Gamma(\nu + k + 1)}$$

$$x^\nu J_\nu(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2\nu+2k}}{2^{\nu+2k}k!\Gamma(\nu + k + 1)}$$

$$\begin{aligned}
\frac{d[x^\nu J_\nu(x)]}{dx} &= \sum_{k=0}^{\infty} \frac{(-1)^k 2(\nu + k)x^{2\nu+2k-1}}{2^{\nu+2k}k!(\nu + k)\Gamma(\nu + k)} \\
&= \sum_{k=0}^{\infty} \frac{(-1)^k x^\nu x^{\nu-1+2k}}{2^{\nu-1+2k}k!\Gamma(\nu - 1 + k + 1)} \\
&= x^\nu \sum_{k=0}^{\infty} \frac{(-1)^k x^{\nu-1+2k}}{2^{\nu-1+2k}k!\Gamma(\nu - 1 + k + 1)} \\
&= x^\nu J_{\nu-1}(x)
\end{aligned}$$

as asserted.

THEOREM 2

$$\frac{d[x^{-\nu}J_\nu(x)]}{dx} = -x^{-\nu}J_{\nu+1}(x)$$

PROOF Theorem 2 can be proved in essentially the same manner as Theorem 1, but it is easier and perhaps more instructive to proceed as follows By performing the indicated differentiations and simplifying, we can verify at once that the differential equation

$$\frac{d}{dx}\left[x^{1-2\nu}\frac{d(x^\nu y)}{dx}\right] + x^{1-\nu}y = 0$$

is precisely Bessel's equation of order ν and is, therefore, satisfied by the particular solution

$$y = J_\nu(x)$$

Hence, substituting, we have

$$\frac{d}{dx}\left\{x^{1-2\nu}\frac{d}{dx}[x^\nu J_\nu(x)]\right\} = -x^{1-\nu}J_\nu(x)$$

Now, using the result of Theorem 1, this can be written

$$\frac{d}{dx}\{x^{1-2\nu}[x^\nu J_{\nu-1}(x)]\} = -x^{1-\nu}J_\nu(x)$$

or

$$\frac{d}{dx}[x^{1-\nu}J_{\nu-1}(x)] = -x^{1-\nu}J_\nu(x)$$

Finally, replacing ν by $\nu + 1$, we have the assertion of Theorem 2

By using their definitions in terms of $J_\nu(x)$ and $J_{-\nu}(x)$, one can readily show that *the Bessel functions of the second kind* $Y_\nu(x)$ *and the Hankel functions* $H_\nu^{(1)}(x)$ *and* $H_\nu^{(2)}(x)$ *also satisfy the identities of Theorems* 1 *and* 2. Furthermore, by arguments similar to those we have just used, the following theorems can be established.

THEOREM 3

$$\frac{d}{dx}[x^\nu I_\nu(x)] = x^\nu I_{\nu-1}(x)$$

THEOREM 4

$$\frac{d}{dx}[x^{-\nu}I_\nu(x)] = x^{-\nu}I_{\nu+1}(x)$$

THEOREM 5

$$\frac{d}{dx}[x^\nu K_\nu(x)] = -x^\nu K_{\nu-1}(x)$$

THEOREM 6

$$\frac{d}{dx}[x^{-\nu}K_\nu(x)] = -x^{-\nu}K_{\nu+1}(x)$$

Performing the differentiations in the identities of Theorems 1 and 2, we obtain, respectively,

$$x^\nu J'_\nu(x) + \nu x^{\nu-1} J_\nu(x) = x^\nu J_{\nu-1}(x)$$

$$x^{-\nu} J'_\nu(x) - \nu x^{-\nu-1} J_\nu(x) = -x^{-\nu} J_{\nu+1}(x)$$

or, dividing the first of these by x^ν and multiplying the second by x^ν and solving for $J'_\nu(x)$ in each case,

(1) $$J'_\nu(x) = J_{\nu-1}(x) - \frac{\nu}{x} J_\nu(x)$$

(2) $$J'_\nu(x) = \frac{\nu}{x} J_\nu(x) - J_{\nu+1}(x)$$

Adding these and dividing by 2, we obtain a third formula for $J'_\nu(x)$

(3) $$J'_\nu(x) = \frac{J_{\nu-1}(x) - J_{\nu+1}(x)}{2}$$

Subtracting (2) from (1) gives the important recurrence formula

$$J_{\nu-1}(x) + J_{\nu+1}(x) = \frac{2\nu}{x} J_\nu(x)$$

Written as

(4) $$J_{\nu+1}(x) = \frac{2\nu}{x} J_\nu(x) - J_{\nu-1}(x)$$

the formula serves to express Bessel functions of higher orders in terms of functions of lower orders, frequently a useful manipulation Written as

(5) $$J_{\nu-1}(x) = \frac{2\nu}{x} J_\nu(x) - J_{\nu+1}(x)$$

it serves similarly to express Bessel functions of large negative orders (for instance) in terms of Bessel functions whose orders are numerically smaller

EXAMPLE 1

Express $J_4(ax)$ in terms of $J_0(ax)$ and $J_1(ax)$

Taking $\nu = 3$ in (4), we first have

$$J_4(ax) = \frac{6}{ax} J_3(ax) - J_2(ax)$$

Applying (4) again to $J_3(ax)$ and then to $J_2(ax)$, we have further

$$\begin{aligned} J_4(ax) &= \frac{6}{ax}\left[\frac{4}{ax} J_2(ax) - J_1(ax)\right] - J_2(ax) \\ &= \left(\frac{24}{a^2x^2} - 1\right) J_2(ax) - \frac{6}{ax} J_1(ax) \\ &= \left(\frac{24}{a^2x^2} - 1\right)\left[\frac{2}{ax} J_1(ax) - J_0(ax)\right] - \frac{6}{ax} J_1(ax) \\ &= \left(\frac{48}{a^3x^3} - \frac{8}{ax}\right) J_1(ax) - \left(\frac{24}{a^2x^2} - 1\right) J_0(ax) \end{aligned}$$

EXAMPLE 2

Show that $$\frac{d[xJ_\nu(x)J_{\nu+1}(x)]}{dx} = x[J_\nu^2(x) - J_{\nu+1}^2(x)]$$

Performing the differentiation, we have

$$\frac{d[xJ_\nu(x)J_{\nu+1}(x)]}{dx} = J_\nu(x)J_{\nu+1}(x) + xJ_\nu'(x)J_{\nu+1}(x) + xJ_\nu(x)J_{\nu+1}'(x)$$

Then, substituting for $xJ_\nu'(x)$ from (2) and for $xJ_{\nu+1}'(x)$ from (1), we have

$$\begin{aligned}\frac{d[xJ_\nu(x)J_{\nu+1}(x)]}{dx} &= J_\nu(x)J_{\nu+1}(x) + J_{\nu+1}(x)[\nu J_\nu(x) - xJ_{\nu+1}(x)] + J_\nu(x)[xJ_\nu(x) - (\nu+1)J_{\nu+1}(x)] \\ &= x[J_\nu^2(x) - J_{\nu+1}^2(x)]\end{aligned}$$

The basic differentiation identities of Theorems 1 and 2, when written as integration formulas

(6) $$\int x^\nu J_{\nu-1}(x)\,dx = x^\nu J_\nu(x) + c$$

(7) $$\int x^{-\nu} J_{\nu+1}(x)\,dx = -x^{-\nu} J_\nu(x) + c$$

suffice for the integration of numerous simple expressions involving Bessel functions For example, taking $\nu = 1$ in (6), we have

$$\int xJ_0(x)\,dx = xJ_1(x) + c$$

Similarly, taking $\nu = 0$ in (7), we find

$$\int J_1(x)\,dx = -J_0(x) + c$$

Usually, however, integration by parts must be used in addition to (6) and (7).

EXAMPLE 3

What is $\int J_3(x)\,dx$?

If we multiply and divide the integrand by x^2, we have $\int x^2[x^{-2}J_3(x)]\,dx$, and so, integrating by parts with

$$u = x^2 \qquad dv = x^{-2}J_3(x)\,dx$$

$$du = 2x\,dx \qquad v = -x^{-2}J_2(x) \qquad \text{[by (7), with } \nu = 2]$$

we have

$$\begin{aligned}\int J_3(x)\,dx &= -J_2(x) + 2\int x^{-1}J_2(x)\,dx \\ &= -J_2(x) - 2x^{-1}J_1(x) + c \qquad \text{[by (7), with } \nu = 1]\end{aligned}$$

EXAMPLE 4

What is $\int [J_2(3x)/x^2]\,dx$?

Here it is convenient to multiply the numerator and denominator of the integrand by $9x^2$, getting

$$\frac{1}{9}\int (3x)^2J_2(3x)\frac{dx}{x^4}$$

Now, integrating by parts with

$$u = (3x)^2J_2(3x) \qquad dv = \frac{dx}{x^4}$$

$$du = (3x)^2J_1(3x)3\,dx \qquad v = -\frac{1}{3x^3}$$

we have
$$\int \frac{J_2(3x)}{x^2}\,dx = \frac{1}{9}\left[-\frac{3J_2(3x)}{x} + 3\int 3xJ_1(3x)\frac{dx}{x^2}\right]$$

Again using integration by parts, with

$$u = 3xJ_1(3x) \qquad dv = \frac{dx}{x^2}$$

$$du = 3xJ_0(3x)3\,dx \qquad v = -\frac{1}{x}$$

we have further

$$\int \frac{J_2(3x)}{x^2}\,dx = \frac{1}{9}\left\{-\frac{3J_2(3x)}{x} + 3\left[-3J_1(3x) + 9\int J_0(3x)\,dx\right]\right\}$$

$$= -\frac{J_2(3x)}{3x} - J_1(3x) + 3\int J_0(3x)\,dx$$

The residual integral $\int J_0(3x)\,dx$ cannot be evaluated in finite form

In general, an integral of the form

$$\int x^m J_n(x)\,dx$$

where m and n are integers such that $m + n \geq 0$, can be completely integrated if $m + n$ is odd, but will ultimately depend upon the residual integral $\int J_0(x)\,dx$ if $m + n$ is even For this reason $\int_0^x J_0(t)\,dt$ has now been tabulated *

Another class of identities of considerable interest can be obtained from the expansion of the function

$$\exp\left[\frac{x}{2}\left(t - \frac{1}{t}\right)\right] = e^{xt/2}e^{-x/2t} \tag{8}$$

in terms of powers of t To derive this expansion we first replace the exponentials on the right of (8) by their infinite series, getting

$$\left(\sum_{i=0}^{\infty} \frac{1}{i!}\frac{x^i t^i}{2^i}\right)\left[\sum_{j=0}^{\infty} \frac{(-1)^j}{j!}\frac{x^j t^{-j}}{2^j}\right]$$

Now when these series are multiplied together, we obtain a term containing t^n $(n \geq 0)$ when and only when the general term in the second series, i e , the term containing t^{-j}, is multiplied by the term in the first series which contains t^{n+j}, i e , the term for which $i = n + j$ Therefore, taking into account all possible values of j, we find that the total coefficient of t^n in the product of the two series is

$$\sum_{j=0}^{\infty}\left[\frac{1}{(n+j)!}\frac{x^{n+j}}{2^{n+j}}\right]\left[\frac{(-1)^j}{j!}\frac{x^j}{2^j}\right] = \sum_{j=0}^{\infty}\frac{(-1)^j x^{n+2j}}{2^{n+2j}j!\,\Gamma(n+j+1)} = J_n(x)$$

Similarly, a term containing t^{-n} arises when and only when the general term in the first series, i e , the term containing t^i, is

---

* A N. Lowan and Milton Abramowitz, "Tables of Integrals of $\int_0^x J_0(t)\,dt$ and $\int_0^x Y_0(t)\,dt$," MT 20, Superintendent of Documents, Government Printing Office, Washington, D.C.

multiplied by the term in the second series which contains t^{-n-i}, i.e., the term for which $j = n + i$. Therefore, taking into account all possible values of i, we find that the total coefficient of t^{-n} in the product of the two series is

$$\sum_{i=0}^{\infty}\left(\frac{1}{i!}\,\frac{x^i}{2^i}\right)\left[\frac{(-1)^{n+i}}{(n+i)!}\,\frac{x^{n+i}}{2^{n+i}}\right] = (-1)^n\sum_{i=0}^{\infty}\frac{(-1)^i x^{n+2i}}{2^{n+2i}\,i!\,\Gamma(n+i+1)}$$

$$= (-1)^n J_n(x)$$

Hence,

$$\exp\left[\frac{x}{2}\left(t-\frac{1}{t}\right)\right] = J_0(x) + \sum_{n=1}^{\infty} J_n(x)[t^n + (-1)^n t^{-n}] \tag{9}$$

Now let $t = e^{i\phi}$, so that

$$\frac{1}{2}\left(t-\frac{1}{t}\right) = \frac{e^{i\phi}-e^{-i\phi}}{2} = i\sin\phi$$

and $$\exp\left[\frac{x}{2}\left(t-\frac{1}{t}\right)\right] = e^{ix\sin\phi} = \cos(x\sin\phi) + i\sin(x\sin\phi)$$

In the same way, when n is even, say $n = 2k$, we have

$$t^n + (-1)^n t^{-n} = t^{2k} + (-1)^{2k}t^{-2k} = e^{i2k\phi} + e^{-i2k\phi} = 2\cos 2k\phi$$

and, when n is odd, say $n = 2k - 1$, we have

$$t^n + (-1)^n t^{-n} = t^{2k-1} + (-1)^{2k-1}t^{-2k+1} = e^{i(2k-1)\phi} - e^{-i(2k-1)\phi}$$

$$= 2i\sin(2k-1)\phi$$

Therefore Eq. (9) can be written

$$e^{ix\sin\phi} \equiv \cos(x\sin\phi) + i\sin(x\sin\phi)$$

$$= J_0(x) + 2\sum_{k=1}^{\infty} J_{2k}(x)\cos 2k\phi + 2i\sum_{k=1}^{\infty} J_{2k-1}(x)\sin(2k-1)\phi$$

Equating real and imaginary parts in the last expression, we obtain the identities

$$\cos(x\sin\phi) = J_0(x) + 2\sum_{k=1}^{\infty} J_{2k}(x)\cos 2k\phi \tag{10}$$

$$\sin(x\sin\phi) = 2\sum_{k=1}^{\infty} J_{2k-1}(x)\sin(2k-1)\phi \tag{11}$$

The series on the right in (10) and (11) are, of course, just the Fourier expansions of the functions on the left.

Now multiply both sides of (10) by $\cos n\phi$ and both sides of (11) by $\sin n\phi$, and integrate each identity with respect to ϕ from 0 to π. Since

$$\int_0^\pi \cos m\phi\cos n\phi\, d\phi = \int_0^\pi \sin m\phi\sin n\phi\, d\phi = 0 \qquad m \neq n$$

$$\int_0^\pi \cos^2 n\phi\, d\phi = \int_0^\pi \sin^2 n\phi\, d\phi = \frac{\pi}{2}$$

this yields

$$\int_0^\pi \cos n\phi \cos (x \sin \phi)\, d\phi = \begin{cases} \pi J_n(x) & n \text{ even} \\ 0 & n \text{ odd} \end{cases}$$

$$\int_0^\pi \sin n\phi \sin (x \sin \phi)\, d\phi = \begin{cases} 0 & n \text{ even} \\ \pi J_n(x) & n \text{ odd} \end{cases}$$

If we add these two expressions and divide by π, we have, for all integral values of n,

$$J_n(x) = \frac{1}{\pi} \int_0^\pi [\cos n\phi \cos (x \sin \phi) + \sin n\phi \sin (x \sin \phi)]\, d\phi$$

since, for every value of n, one or the other of the integrals vanishes while the remaining one contributes $J_n(x)$ Finally, using the formula for the cosine of the difference of two quantities, we have

$$J_n(x) = \frac{1}{\pi} \int_0^\pi \cos (n\phi - x \sin \phi)\, d\phi \qquad n \text{ an integer} \tag{12}$$

EXERCISES

1 Express $J_5(x)$ in terms of $J_0(x)$ and $J_1(x)$

2 Express $J_{3/2}(x)$ and $J_{-3/2}(x)$ in terms of $\sin x$ and $\cos x$

3 What is $\frac{d[x^2J_3(2x)]}{dx}$?

4 What is $\frac{d[xJ_0(x^2)]}{dx}$?

5 Show that $\frac{d[x^2J_\nu(x)J_{\nu+1}(x)]}{dx} = 2x^2J_\nu(x)\frac{dJ_\nu(x)}{dx}$

6 Prove Theorem 2 by using the series expansion for $J_\nu(x)$

7 Show that

a $4J_\nu''(x) = J_{\nu-2}(x) - 2J_\nu(x) + J_{\nu+2}(x)$

b $8J_\nu'''(x) = J_{\nu-3}(x) - 3J_{\nu-1}(x) + 3J_{\nu+1}(x) - J_{\nu+3}(x)$

8 Show that $J_\nu''(x) = \left[\frac{\nu(\nu + 1)}{x^2} - 1\right] J_\nu(x) - \frac{J_{\nu-1}(x)}{x}$

9 Show that $J_0(x) = \frac{1}{\pi}\int_0^\pi \cos (x \cos \phi)\, d\phi$

10 By expanding the integrand into an infinite series and integrating term by term, show that

a $\int_0^{\pi/2} J_0(x \cos \phi) \cos \phi\, d\phi = \frac{\sin x}{x}$ b $\int_0^{\pi/2} J_1(x \cos \phi)\, d\phi = \frac{1 - \cos x}{x}$

11 Show that $\int J_0(x)\, dx = 2[J_1(x) + J_3(x) + J_5(x) + \qquad]$ [Hint· Use Formula (3)]

12 Show that

$$\begin{aligned}\int J_0(x)\, dx &= J_1(x) + \int \frac{J_1(x)}{x}\, dx \\ &= J_1(x) + \frac{J_2(x)}{x} + 1\ \ 3 \int \frac{J_2(x)}{x^2}\, dx \\ &= J_1(x) + \frac{J_2(x)}{x} + \frac{1\ \ 3}{x^2} J_3(x) + 1\ \ 3\ \ 5 \int \frac{J_3(x)}{x^3}\, dx \\ &\quad \cdot\ \cdot \qquad \cdot\ \cdot\ \cdot \qquad \cdot \qquad \cdot \qquad \cdot \\ &= J_1(x) + \frac{J_2(x)}{x} + \frac{1\ \ 3}{x^2} J_3(x) + \qquad + \frac{(2n-2)!J_n(x)}{2^{n-1}(n-1)!x^{n-1}} + \frac{(2n)!}{2^n n!} \int \frac{J_n(x)}{x^n}\, dx\end{aligned}$$

[Hint· Use repeated integration by parts, each time taking $dv = x^{k+1}J_k(x)\, dx$]

13 Show that $\int xJ_m^2(x)\,dx = x^2\left[\frac{J_m^2(x) - J_{m-1}(x)J_{m+1}(x)}{2}\right] + c$

(Hint After integrating by parts, the result of Exercise 5 may be helpful.)

14 Show that $\int J_0(x)\cos x\,dx = xJ_0(x)\cos x + xJ_1(x)\sin x + c$

15 Show that $\int J_0(x)\sin x\,dx = xJ_0(x)\sin x - xJ_1(x)\cos x + c$

16 Show that $\int J_1(x)\cos x\,dx = xJ_1(x)\cos x - J_0(x)(x\sin x + \cos x) + c$

17 Show that $\int J_1(x)\sin x\,dx = xJ_1(x)\sin x + J_0(x)(x\cos x - \sin x) + c$

18 What is

a $\int xJ_0(x)\cos x\,dx$? **b** $\int xJ_1(x)\sin x\,dx$?

19 What is

a $\int xJ_0(x)\sin x\,dx$? **b** $\int xJ_1(x)\cos x\,dx$?

20 Show that

a $\int xJ_0(x)\,dx = xJ_1(x) + c$

b $\int x^2J_0(x)\,dx = x^2J_1(x) + xJ_0(x) - \int J_0(x)\,dx + c$

c $\int x^3J_0(x)\,dx = (x^3 - 4x)J_1(x) + 2x^2J_0(x) + c$

d $\int x^4J_0(x)\,dx = (x^4 - 9x^2)J_1(x) + (3x^3 - 9x)J_0(x) + 9\int J_0(x)\,dx + c$

21 Show that

a $\displaystyle\int \frac{J_1(x)}{x}\,dx = -J_1(x) + \int J_0(x)\,dx + c$

b $\int J_1(x)\,dx = -J_0(x) + c$

c $\int xJ_1(x)\,dx = -xJ_0(x) + \int J_0(x)\,dx + c$

d $\int x^2J_1(x)\,dx = 2xJ_1(x) - x^2J_0(x) + c$

e $\int x^3J_1(x)\,dx = 3x^2J_1(x) - (x^3 - 3x)J_0(x) - 3\int J_0(x)\,dx + c$

f $\int x^4J_1(x)\,dx = (4x^3 - 16x)J_1(x) - (x^4 - 8x^2)J_0(x) + c$

22 What is $\int xJ_2(1 - x)\,dx$? **23** What is $\int J_0(\sqrt{x})\,dx$?

24 Show that

a $I'_\nu(x) = I_{\nu-1}(x) - \frac{\nu}{x}I_\nu(x)$ **b** $I'_\nu(x) = \frac{\nu}{x}I_\nu(x) + I_{\nu+1}(x)$

c $I'_\nu(x) = \frac{I_{\nu-1}(x) + I_{\nu+1}(x)}{2}$ **d** $I_{\nu-1}(x) - I_{\nu+1}(x) = \frac{2\nu}{x}I_\nu(x)$

25 What is

a $\int xI_0(x)dx$? **b** $\int x^2I_0(x)dx$?

c $\int xI_1(x)\,dx$? **d** $\int x^2I_1(x)\,dx$?

9.6

Orthogonality of the Bessel functions

If we write Bessel's equation of order ν in the form

$$x\frac{d^2y}{dx^2} + \frac{dy}{dx} + \left(\lambda^2x - \frac{\nu^2}{x}\right)y \equiv \frac{d(xy')}{dx} + \left(-\frac{\nu^2}{x} + \lambda^2x\right)y = 0$$

it is clear that it is a special case, with

$$p(x) = x \qquad q(x) = -\frac{\nu^2}{x} \qquad r(x) = x$$

and λ^2 written in place of λ, of the general equation covered by Theorem 4, Sec 8 5 If the solutions of Bessel's equation satisfy boundary conditions of the form

$$(1) \qquad A_iy_\nu(\lambda x_i) - B_i\frac{dy_\nu(\lambda x)}{dx}\bigg|_{x=x_i} = 0 \qquad i = 1, 2$$

they must, therefore, be orthogonal with respect to the weight function $p(x) = x$ over the interval (x_1,x_2).†

For practical purposes, however, it is not enough to know that the characteristic functions of a problem are orthogonal. In order to carry out the expansions required at the final stage of a typical boundary value problem, it is also necessary to know the value of the integral of the product of the weight function and the square of the general characteristic function, taken over the interval of the problem

We begin this calculation by considering the indefinite integral $\int t y_\nu^2(t)\,dt$ where $y_\nu(t)$ is *any* solution of Bessel's equation, i e ,

$$t^2 y_\nu'' + t y_\nu' + (t^2 - \nu^2) y_\nu = 0 \tag{2}$$

If Eq (2) is multiplied by y_ν' and then integrated, we obtain

$$\int t^2 y_\nu' y_\nu''\,dt + \int t (y_\nu')^2\,dt + \int t^2 y_\nu y_\nu'\,dt - \nu^2 \int y_\nu y_\nu'\,dt = 0 \tag{3}$$

Now, evaluating the first and third integrals by parts, we have

$$\int t^2 y_\nu' y_\nu''\,dt \xrightarrow[\substack{u = t^2 \qquad dv = y_\nu' y_\nu''\,dt \\ du = 2t\,dt \qquad v = \frac12 (y_\nu')^2}]{} \tfrac12 t^2 (y_\nu')^2 - \int t (y_\nu')^2\,dt$$

$$\int t^2 y_\nu y_\nu'\,dt \xrightarrow[\substack{u = t^2 \qquad dv = y_\nu y_\nu'\,dt \\ du = 2t\,dt \qquad v = \frac12 y_\nu^2}]{} \tfrac12 t^2 y_\nu^2 - \int t y_\nu^2\,dt$$

Then, substituting these results into Eq (3), we find

$$\left[\tfrac12 t^2 (y_\nu')^2 - \int t (y_\nu')^2\,dt\right] + \int t (y_\nu')^2\,dt + \left(\tfrac12 t^2 y_\nu^2 - \int t y_\nu^2\,dt\right) - \tfrac12 \nu^2 y_\nu^2 = 0$$

or, collecting terms and solving for $\int t y_\nu^2\,dt$,

$$\int t y_\nu^2(t)\,dt = \frac12 (t^2 - \nu^2) y_\nu^2(t) + \frac12 t^2 \left[\frac{dy_\nu(t)}{dt}\right]^2$$

If we now put $t = \lambda_m x$, where λ_m is any one of the characteristic values for which solutions satisfying the boundary conditions exist, and then divide by λ_m^2, we obtain the integral in which we are actually interested

$$\int x y_\nu^2(\lambda_m x)\,dx = \frac{1}{2\lambda_m^2} \left\{ (\lambda_m^2 x^2 - \nu^2) y_\nu^2(\lambda_m x) + x^2 \left[\frac{dy_\nu(\lambda_m x)}{dx}\right]^2 \right\} \tag{4}$$

The evaluation of (4) between the specific limits x_1 and x_2 requires the consideration of several special cases, according as B_i in the boundary conditions (1) is or is not equal to zero. If $B_i = 0$, then (1) becomes simply

$$y_\nu(\lambda_m x_i) = 0 \tag{5}$$

and the antiderivative on the right of (4) reduces to

$$\frac{1}{2\lambda_m^2} x_i^2 \left[\frac{dy_\nu(\lambda_m x)}{dx}\right]^2 \Bigg|_{x = x_i} \tag{6}$$

---

† Since $r(x) \equiv x$ vanishes when $x = 0$, it follows from the proof of Theorem 4, Sec 8.5, that if $x_1 = 0$, no boundary condition will be needed (and none will be available) at $x = x_1$

This can be further simplified by recalling from the preceding section that all solutions of Bessel's equation J_ν, $J_{-\nu}$, Y_ν, $H_\nu^{(1)}$, and $H_\nu^{(2)}$, as well as arbitrary linear combinations of these functions, satisfy the identity

$$t\frac{dy_\nu(t)}{dt} = \nu y_\nu(t) - t y_{\nu+1}(t)$$

or
$$x\frac{dy_\nu(\lambda_m x)}{dx} = \nu y_\nu(\lambda_m x) - \lambda_m x\, y_{\nu+1}(\lambda_m x)$$

Evaluating this at $x = x_i$ and using (5), we find

$$x_i\frac{dy_\nu(\lambda_m x)}{dx}\bigg|_{x=x_i} = -\lambda_m x_i\, y_{\nu+1}(\lambda_m x_i)$$

and so (6) becomes simply

$$\tfrac{1}{2}x_i^2 y_{\nu+1}^2(\lambda_m x_i)$$

On the other hand, if $B_i \neq 0$, then we can substitute for the derivative on the right of Eq. (4), getting

$$\frac{y_\nu^2(\lambda_m x_i)}{2\lambda_m^2}\left[(\lambda_m x_i)^2 - \nu^2 + \left(\frac{x_i A_i}{B_i}\right)^2\right]$$

The results of the preceding discussion are summarized in the following important theorem.

THEOREM 1

The solutions of Bessel's equation of order ν which satisfy the boundary conditions

$$A_i y_\nu(\lambda x_i) - B_i\frac{dy_\nu(\lambda x)}{dx}\bigg|_{x=x_i} = 0 \qquad i = 1, 2$$

form an orthogonal system with respect to the weight function x over the interval (x_1,x_2). The integral of the product of the weight function and the square of any solution of the system $\{y_\nu(\lambda_m x)\}$, i.e.,

$$\int_{x_1}^{x_2} x y_\nu^2(\lambda_m x)\, dx$$

is equal to

$$\frac{y_\nu^2(\lambda_m x_2)}{2\lambda_m^2}\left[(\lambda_m x_2)^2 - \nu^2 + \left(\frac{x_2 A_2}{B_2}\right)^2\right] - \frac{y_\nu^2(\lambda_m x_1)}{2\lambda_m^2}\left[(\lambda_m x_1)^2 - \nu^2 + \left(\frac{x_1 A_1}{B_1}\right)^2\right] \qquad B_1B_2 \neq 0$$

$$\frac{y_\nu^2(\lambda_m x_2)}{2\lambda_m^2}\left[(\lambda_m x_2)^2 - \nu^2 + \left(\frac{x_2 A_2}{B_2}\right)^2\right] - \frac{x_1^2}{2}y_{\nu+1}^2(\lambda_m x_1) \qquad B_1 = 0,\ B_2 \neq 0$$

$$\frac{x_2^2}{2}y_{\nu+1}^2(\lambda_m x_2) - \frac{y_\nu^2(\lambda_m x_1)}{2\lambda_m^2}\left[(\lambda_m x_1)^2 - \nu^2 + \left(\frac{x_1 A_1}{B_1}\right)^2\right] \qquad B_1 \neq 0,\ B_2 = 0$$

$$\frac{x_2^2}{2}y_{\nu+1}^2(\lambda_m x_2) - \frac{x_1^2}{2}y_{\nu+1}^2(\lambda_m x_1) \qquad B_1 = B_2 = 0$$

If $x_1 = 0$, no boundary condition is needed at $x = x_1$, and the contribution to the integral from the lower limit is zero.

EXAMPLE 1

Expand $f(x) = 4x - x^3$ over the interval (0,2) in terms of the Bessel functions of the first kind of order 1 which satisfy the boundary condition

$$J_1(\lambda x)\Big|_{x=2} = 0$$

In this case the characteristic values are the values of λ determined by the roots of the equation

$$J_1(2\lambda) = 0$$

Now the roots of the equation $J_1(z) = 0$ are*

$$z_0 = 0 \qquad z_1 = 3.832 \qquad z_2 = 7.016 \qquad z_3 = 10.174 \qquad z_4 = 13.324$$

Hence, $\qquad \lambda_0 = 0 \qquad \lambda_1 = 1.916 \qquad \lambda_2 = 3.508 \qquad \lambda_3 = 5.087 \qquad \lambda_4 = 6.662$

Therefore, since $J_1(\lambda_0 x) \equiv J_1(0) = 0$, the characteristic functions in terms of which the expansion is to be carried out are

$$J_1(\lambda_1 x) \qquad J_1(\lambda_2 x) \qquad J_1(\lambda_3 x) \qquad J_1(\lambda_4 x)$$

As in the simpler case of Fourier expansions, we begin by writing

$$f(x) = 4x - x^3 = A_1 J_1(\lambda_1 x) + A_2 J_1(\lambda_2 x) + \cdots + A_m J_1(\lambda_m x) + \cdots$$

Multiplying both sides of this expression by $xJ_1(\lambda_m x)$, integrating from 0 to 2, and using the results of Theorem 1, we have

$$\int_0^2 (4x - x^3)xJ_1(\lambda_m x)\,dx = A_m \int_0^2 xJ_1^2(\lambda_m x)\,dx = 2A_m J_2^2(2\lambda_m)$$

Hence
$$A_m = \frac{\displaystyle\int_0^2 (4x^2 - x^4)J_1(\lambda_m x)\,dx}{2J_2^2(2\lambda_m)}$$

For the integral

$$4\int_0^2 x^2 J_1(\lambda_m x)\,dx = \frac{4}{\lambda_m^3}\int_0^2 (\lambda_m x)^2 J_1(\lambda_m x)\,d(\lambda_m x)$$

we have immediately, from Eq. (6), Sec. 9.5,

$$\frac{4}{\lambda_m^3}(\lambda_m x)^2 J_2(\lambda_m x)\Big|_0^2 = \frac{16}{\lambda_m} J_2(2\lambda_m)$$

To evaluate
$$\int_0^2 x^4 J_1(\lambda_m x)\,dx = \frac{1}{\lambda_m^5}\int_0^2 (\lambda_m x)^4 J_1(\lambda_m x)\,d(\lambda_m x) = \frac{1}{\lambda_m^5}\int_0^{2\lambda_m} t^4 J_1(t)\,dt$$

we use integration by parts, with

$$u = t^2 \qquad dv = t^2 J_1(t)\,dt$$
$$du = 2t\,dt \qquad v = t^2 J_2(t)$$

This gives
$$\int_0^2 x^4 J_1(\lambda_m x)\,dx = \frac{1}{\lambda_m^5}\left[t^4 J_2(t)\Big|_0^{2\lambda_m} - 2\int_0^{2\lambda_m} t^3 J_2(t)\,dt \right]$$
$$= \frac{1}{\lambda_m^5}[t^4 J_2(t) - 2t^3 J_3(t)]_0^{2\lambda_m}$$
$$= \frac{16}{\lambda_m^2}[\lambda_m J_2(2\lambda_m) - J_3(2\lambda_m)]$$

---

* See, for instance, Eugene Jahnke, Fritz Emde, and Friedrich Losch, "Tables of Higher Functions," 6th ed., p. 193, McGraw-Hill Book Company, New York, 1960.

Thus,
$$A_m = \frac{1}{2J_2^2(2\lambda_m)}\left\{\frac{16}{\lambda_m}J_2(2\lambda_m) - \frac{16}{\lambda_m^2}[\lambda_m J_2(2\lambda_m) - J_3(2\lambda_m)]\right\} = \frac{8J_3(2\lambda_m)}{\lambda_m^2 J_2^2(2\lambda_m)}$$

But, by Formula (4), Sec 9 5,

$$J_3(2\lambda_m) = \frac{4}{2\lambda_m}J_2(2\lambda_m) - J_1(2\lambda_m) = \frac{2J_2(2\lambda_m)}{\lambda_m}$$

since the λ's were determined by the condition that $J_1(2\lambda_m) = 0$. Therefore A_m can be further simplified to

$$A_m = \frac{16}{\lambda_m^3 J_2(2\lambda_m)}$$

The same reduction can be repeated for $J_2(2\lambda_m)$, since

$$J_2(2\lambda_m) = \frac{2}{2\lambda_m}J_1(2\lambda_m) - J_0(2\lambda_m) = -J_0(2\lambda_m)$$

Hence, finally, $$A_m = -\frac{16}{\lambda_m^3 J_0(2\lambda_m)}$$

The required expansion is, therefore,

$$4x - x^3 = -16\sum_{m=1}^{\infty}\frac{J_1(\lambda_m x)}{\lambda_m^3 J_0(2\lambda_m)}$$

Plots showing the degree to which the first term and the first two terms of this series approximate $4x - x^3$ are shown in Fig 9 7

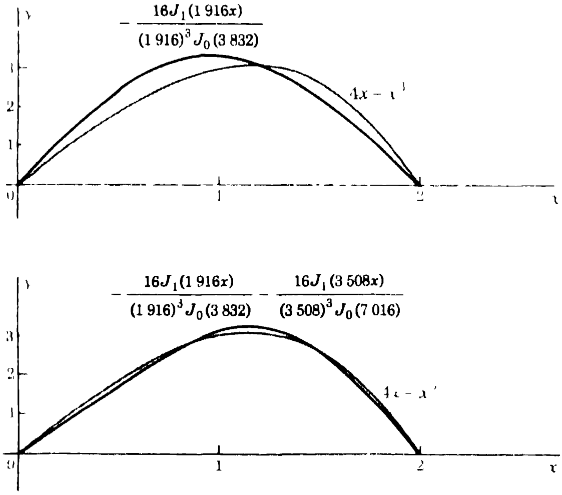


FIGURE 9.7
Plot showing the approximation of a function by the first two terms of a Bessel function expansion.

EXERCISES

1 Expand $f(x) = 1$ over the interval (0,3) in terms of the functions $J_0(\lambda_m x)$, where the λ's are determined by $J_0(3\lambda) = 0$.

2 Expand $f(x) = 1$ over the interval (0,3) in terms of the functions $J_2(\lambda_m x)$, where the λ's are determined by $J_2(3\lambda) = 0$

3 Expand $f(x) = x$ over the interval $(0,2)$ in terms of the functions $J_1(\lambda_m x)$, where the λ's are determined by $J_1(2\lambda) = 0$

4 Expand $f(x) = x^2$ over the interval $(0,3)$ in terms of the functions $J_0(\lambda_m x)$ where the λ's are determined by $\left.\frac{dJ_0(\lambda x)}{dx}\right|_{x=3} = 0$

5 Expand $f(x) = x^2$ over the interval $(0,1)$ in terms of the functions $J_2(\lambda_m x)$, where the λ's are determined by $J_2(\lambda) = 0$

6 Expand

$$f(x) = \begin{cases} x & 0 < x < 1 \\ 0 & 1 < x < 2 \end{cases}$$

in terms of the functions $J_1(\lambda_m x)$, where the λ's are determined by

$$\left.\frac{dJ_1(\lambda x)}{dx}\right|_{x=2} = 0$$

7 Expand $f(x) = 1$ over the interval $(0,3)$ in terms of the functions $J_0(\lambda_m x)$, where the λ's are determined by

$$J_0(3\lambda) - \left.\frac{dJ_0(\lambda x)}{dx}\right|_{x=3} = 0$$

8 Using tables of the Bessel functions, compute the first two characteristic values in Exercise 7 correct to two decimal places

9 If the boundary conditions in Theorem 1 are of the form $y_\nu(\lambda x) = 0$ at $x = 1$ and at $x = 5$, what is the equation satisfied by the characteristic values $\{\lambda_m\}$?

10 Does Theorem 1 have a counterpart for the modified Bessel equation? Why?

9.7

Applications of Bessel functions

Bessel functions occur in a great many practical problems. In principle they are always to be expected when partial differential equations are applied to configurations possessing circular symmetry. On the other hand, they also arise in numerous applications where neither circular symmetry nor partial differential equations are involved In this section we shall conclude our treatment of Bessel functions by discussing a variety of problems where their use is required

EXAMPLE 1

What is $\mathcal{L}\{t^\nu J_\nu(\lambda t)\}$ if $\nu \geqq 0$?

It is possible to determine the required transform by expressing $t^\nu J_\nu(\lambda t)$ as an infinite series and then taking the transform term by term However, it is more instructive to proceed as follows:

From Corollary 1 of Theorem 1, Sec 9 4, it is clear that $y = t^\nu J_\nu(\lambda t)$ is a solution of the differential equation

$$(t^{1-2\nu}y')' + \lambda^2 t^{1-2\nu}y = 0$$

that is,

$$ty'' + (1 - 2\nu)y' + \lambda^2 ty = 0$$

If we take the Laplace transform of this equation, recalling Theorem 7, Sec. 7.4, we obtain

$$-\frac{d}{ds}(s^2\mathcal{L}\{y\} - sy_0 - y_0') + (1 - 2\nu)(s\mathcal{L}\{y\} - y_0) - \lambda^2\frac{d}{ds}\mathcal{L}\{y\}$$

$$= -s^2\frac{d\mathcal{L}\{y\}}{ds} - 2s\mathcal{L}\{y\} + y_0 + (1 - 2\nu)(s\mathcal{L}\{y\} - y_0) - \lambda^2\frac{d\mathcal{L}\{y\}}{ds}$$

$$= -(s^2 + \lambda^2)\frac{d\mathcal{L}\{y\}}{ds} - (1 + 2\nu)s\mathcal{L}\{y\} + 2\nu y_0 = 0$$

Now, if $\nu \geqq 0$, the term $2\nu y_0$ vanishes identically, because either $\nu = 0$ or else

$$y_0 = t^\nu J_\nu(\lambda t)\Big|_{t=0} = 0$$

Hence, the last equation reduces to the separable differential equation

$$\frac{d\mathcal{L}\{y\}}{\mathcal{L}\{y\}} + (1 + 2\nu)\frac{s\,ds}{s^2 + \lambda^2} = 0$$

Integrating this, we have

$$\ln \mathcal{L}\{y\} + \frac{1 + 2\nu}{2}\ln(s^2 + \lambda^2) = \ln c$$

and, therefore, $\quad \mathcal{L}\{y\} \equiv \mathcal{L}\{t^\nu J_\nu(\lambda t)\} = \dfrac{c}{(s^2 + \lambda^2)^{(1+2\nu)/2}}$

To determine c we consider the leading term on each side of the last equality.

$$\mathcal{L}\left\{t^\nu\left(\frac{\lambda^\nu t^\nu}{2^\nu\Gamma(\nu + 1)} - \qquad\right)\right\} = \frac{c}{(s^2 + \lambda^2)^{(1+2\nu)/2}}$$

$$\frac{\lambda^\nu}{2^\nu\Gamma(\nu + 1)}\mathcal{L}\{t^{2\nu} - \qquad\} = \frac{c}{s^{2\nu+1}}\left(1 + \frac{\lambda^2}{s^2}\right)^{-(2\nu+1)/2}$$

$$\frac{\lambda^\nu}{2^\nu\Gamma(\nu + 1)}\left[\frac{\Gamma(2\nu + 1)}{s^{2\nu+1}} - \qquad\right] = \frac{c}{s^{2\nu+1}}(1 - \qquad)$$

Hence, since this must be an identity, we find

$$c = \frac{\lambda^\nu\Gamma(2\nu + 1)}{2^\nu\Gamma(\nu + 1)}$$

and so

$$\mathcal{L}\{t^\nu J_\nu(\lambda t)\} = \frac{\lambda^\nu\Gamma(2\nu + 1)}{2^\nu\Gamma(\nu + 1)(s^2 + \lambda^2)^{(2\nu+1)/2}} \qquad \nu \geqq 0 \tag{1}$$

Numerous other transform formulas can be obtained from (1). For instance, since, from (1),

$$\mathcal{L}\{J_0(\lambda t)\} = \frac{1}{\sqrt{s^2 + \lambda^2}} \qquad \text{and} \qquad \frac{dJ_0(\lambda t)}{dt} = -\lambda J_1(\lambda t)$$

it follows that

$$\mathcal{L}\{J_1(\lambda t)\} = -\frac{1}{\lambda}\mathcal{L}\left\{\frac{dJ_0(\lambda t)}{dt}\right\} = -\frac{1}{\lambda}[s\mathcal{L}\{J_0(\lambda t)\} - J_0(0)]$$

$$= -\frac{1}{\lambda}\left(\frac{s}{\sqrt{s^2 + \lambda^2}} - 1\right) = \frac{1}{\lambda}\left(\frac{\sqrt{s^2 + \lambda^2} - s}{\sqrt{s^2 + \lambda^2}}\right)$$

$$= \frac{\lambda}{\sqrt{s^2 + \lambda^2}\,(s + \sqrt{s^2 + \lambda^2})}$$

Other results will be found among the exercises.

EXAMPLE 2

A uniform, perfectly flexible cable of length l and weight per unit length w hangs by one end from a frictionless hook. At $t = 0$, while the cable is at rest in a vertical position, a uniform horizontal velocity v is imparted to the portion of the cable between $x = 0$ and $x = \alpha l$ (Fig. 9.8). Find the expression describing the subsequent motion of the cable.

This is essentially the problem of the vibrating string discussed in Sec. 8.2 except for one important difference. Here, instead of being constant, the tension at a general point of the cable is equal to the weight wx of the portion of the cable below that point. Hence, in this case Eq. (1), Sec. 8.2, becomes in the limit

$$\frac{w}{g}\frac{\partial^2 y}{\partial t^2} = \frac{\partial\left(wx\dfrac{\partial y}{\partial x}\right)}{\partial x}$$

As usual, we assume a product solution $y = X(x)T(t)$ and attempt to separate variables. Then, substituting, we have

$$T''X = gT(xX')' \qquad \text{or} \qquad \frac{(xX')'}{X} = \frac{T''}{gT}$$

The common value of these two fractions must be a negative constant, say $-\lambda^2$, for otherwise T will not be a periodic function, as we know it must be. Hence,

$$T = A\cos\lambda\sqrt{g}\,t + B\sin\lambda\sqrt{g}\,t$$

and

$$(2) \qquad (xX')' + \lambda^2 X = 0$$

Using Corollary 1 of Theorem 1, Sec. 9.4, the solution for X is found at once to be

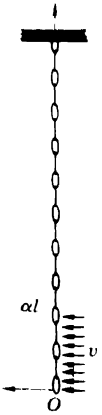


FIGURE 9.8 A hanging cable acted upon by a transverse force over a portion of its length.

$$X = CJ_0(2\lambda\sqrt{x}) + DY_0(2\lambda\sqrt{x})$$

Since the displacement of the free end of the cable will obviously be finite, whereas $Y_0(2\lambda\sqrt{x})$ becomes infinite as x approaches zero, it is clear that D must be zero. Moreover, for all values of t, y is zero when $x = l$. Hence $X(l) = 0$, that is,

$$(3) \qquad J_0(2\lambda\sqrt{l}) = 0$$

This, of course, is the frequency equation of the system. It has infinitely many roots,

$$2\lambda\sqrt{l} = 2.4048, \qquad 5.5201, \qquad 8.6537, \qquad \ldots$$

and so the natural frequencies of the cable, namely, $\omega_n = \lambda_n\sqrt{g}$, are

$$\omega_1 = 1.2024\sqrt{g/l}, \qquad \omega_2 = 2.7600\sqrt{g/l}, \qquad \omega_3 = 4.3268\sqrt{g/l}, \qquad \ldots$$

We have now been led to an infinite sequence of product solutions,

$$y_m(x,t) = X_m(x)T_m(t) = J_0(2\lambda_m\sqrt{x})[A_m\cos(\lambda_m\sqrt{g}\,t) + B_m\sin(\lambda_m\sqrt{g}\,t)]$$

None of these by itself can satisfy the given initial conditions, namely,

$$y(x,0) = 0$$

$$\left.\frac{\partial y}{\partial t}\right|_{x,0} \equiv f(x) = \begin{cases} v & 0 < x < \alpha l \\ 0 & \alpha l < x < l \end{cases}$$

Hence, as usual, we form an infinite series of the individual product solutions,

$$(4) \qquad y(x,t) = \sum_{m=1}^{\infty} J_0(2\lambda_m \sqrt{x})[A_m \cos(\lambda_m \sqrt{g}\, t) + B_m \sin(\lambda_m \sqrt{g}\, t)]$$

and attempt to make it fit the initial conditions.

Now Eq (2) with its accompanying boundary condition $X(l) = 0$ meets all the conditions of Theorem 4, Sec. 8 5 Hence, the X's are orthogonal with respect to the weight function $p(x) \equiv 1$ over the interval $(0,l)$, and thus the A's and B's can be determined by the familiar generalized Fourier procedure To find A_m we put $t = 0$ and $y = 0$ in (4), getting

$$0 = \sum_{m=1}^{\infty} A_m J_0(2\lambda_m \sqrt{x})$$

from which it is obvious that

$$A_m = 0 \qquad m = 1, 2, 3, \ldots$$

To find B_m we differentiate (4) with respect to t and then put $t = 0$ and $\left.\frac{\partial y}{\partial t}\right|_{t=0} = f(x)$, getting

$$(5) \qquad f(x) = \sum_{m=1}^{\infty} \sqrt{g}\, \lambda_m B_m J_0(2\lambda_m \sqrt{x})$$

Next, we multiply (4) by $J_0(2\lambda_m \sqrt{x})$ and integrate from 0 to l From the orthogonality of the J_0's, every term on the right but one becomes zero, and we have

$$\int_0^l f(x) J_0(2\lambda_m \sqrt{x})\, dx \equiv \int_0^{\alpha l} v J_0(2\lambda_m \sqrt{x})\, dx = \sqrt{g}\, \lambda_m B_m \int_0^l J_0^2(2\lambda_m \sqrt{x})\, dx$$

or

$$B_m = \frac{v \int_0^{\alpha l} J_0(2\lambda_m \sqrt{x})\, dx}{\sqrt{g}\, \lambda_m \int_0^l J_0^2(2\lambda_m \sqrt{x})\, dx}$$

To evaluate these integrals we make the obvious substitutions $x = u^2$ and $dx = 2u\, du$,

getting

$$B_m = \frac{v \int_0^{\sqrt{\alpha l}} u J_0(2\lambda_m u)\, du}{\sqrt{g}\, \lambda_m \int_0^{\sqrt{l}} u J_0^2(2\lambda_m u)\, du}$$

The integral in the numerator is precisely

$$\left.\frac{u J_1(2\lambda_m u)}{2\lambda_m}\right|_0^{\sqrt{\alpha l}} = \frac{\sqrt{\alpha l}\, J_1(2\lambda_m \sqrt{\alpha l})}{2\lambda_m}$$

Because of the condition (3), the value of the integral in the denominator is, as we showed in the proof of Theorem 1, Sec 9 6,

$$\frac{l J_1^2(2\lambda_m \sqrt{l})}{2}$$

Hence, finally,

$$B_m = \frac{v}{\lambda_m^2} \sqrt{\frac{\alpha}{gl}}\, \frac{J_1(2\lambda_m \sqrt{\alpha l})}{J_1^2(2\lambda_m \sqrt{l})}$$

With A_m and B_m determined for all values of m, the solution is now complete

It is interesting to note that, since the λ's are incommensurable, there are no two times when the terms $\sin \sqrt{g}\, \lambda_m t$ are respectively the same Hence the cable never returns to a position coinciding exactly with an earlier one unless it is vibrating in one of its normal modes, that is, unless all but one of the B_m's are zero, which cannot happen for the given $f(x)$ This is in sharp contrast to the behavior of the string stretched under uniform tension, which repeats *any* configuration exactly after intervals of $2l/a$, where a is the propagation velocity for the string.

EXAMPLE 3

A metal fin of triangular cross section is attached to a plane surface to help carry off heat from the latter. Assuming dimensions and coordinates as shown in Fig 9 9, find the steady-state temperature distribution along the fin if the wall temperature is u_w and if the fin cools freely into air of constant temperature u_0

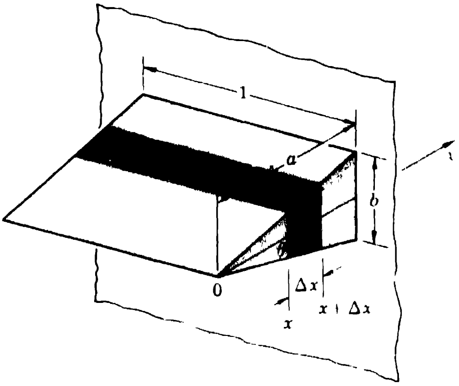


FIGURE 9.9 A portion of a triangular cooling fin attached to a flat wall

We shall base our analysis upon a unit length of the fin and shall assume that the fin is so thin that temperature variations parallel to the base can be neglected Now, consider the heat balance in the element of the fin between x and $x + \Delta x$ This element gains heat by internal flow through its right face and loses heat by internal flow through its left face and also by cooling through its upper and lower surfaces Through the right face the gain of heat per unit time is

Area × thermal conductivity × temperature gradient

or
$$\left[\left(1 \cdot \frac{bx}{a}\right) k \frac{du}{dx}\right]_{x+\Delta x} = \left[\frac{bkx}{a}\frac{du}{dx}\right]_{x+\Delta x}$$

Through the left face the element loses heat at the rate

$$\left[\frac{bkx}{a}\frac{du}{dx}\right]_x$$

Through the surfaces exposed to the air the element loses heat at the rate

Area × surface conductivity × (surface temperature − air temperature)

or
$$2\left(1 \cdot \frac{\Delta x}{\cos\theta}\right) h(u - u_0) = \frac{2h(u - u_0)\,\Delta x}{\cos\theta}$$

Under steady-state conditions the rate of gain of heat must equal the rate of loss, and thus we have

$$\left[\frac{bkx}{a}\frac{du}{dx}\right]_{x+\Delta x} = \left[\frac{bkx}{a}\frac{du}{dx}\right]_x + \frac{2h(u - u_0)\,\Delta x}{\cos\theta}$$

Writing this as

$$\frac{[x(du/dx)]_{x+\Delta x} - [x(du/dx)]_x}{\Delta x} - \frac{2ah}{bk\cos\theta}(u - u_0) = 0$$

and letting $\Delta x \to 0$, we obtain the differential equation

$$\frac{d(xu')}{dx} - \frac{2ah}{bk\cos\theta}(u - u_0) = 0$$

If we set $\qquad U = u - u_0 \qquad$ and $\qquad \alpha^2 = \dfrac{2ah}{bk\cos\theta}$

this becomes
$$\frac{d(xU')}{dx} - \alpha^2 U = 0$$

This can be solved immediately by means of the corollary of Theorem 1, Sec 9.4, and we have

$$U = u - u_0 = c_1 I_0(2\alpha\sqrt{x}) + c_2 K_0(2\alpha\sqrt{x})$$

Since $K_0(2\alpha\sqrt{x})$ is infinite when $x = 0$, c_2 must be zero, leaving

$$u - u_0 = c_1 I_0(2\alpha\sqrt{x})$$

Furthermore, $u = u_w$ when $x = a$, hence,

$$u_w - u_0 = c_1 I_0(2\alpha\sqrt{a}) \qquad \text{or} \qquad c_1 = \frac{u_w - u_0}{I_0(2\alpha\sqrt{a})}$$

Therefore,
$$u = u_0 + (u_w - u_0)\frac{I_0(2\alpha\sqrt{x})}{I_0(2\alpha\sqrt{a})}$$

EXAMPLE 4

A solid consists of one-half of a right circular cylinder of radius b and height h (Fig 9 10) The lower base, the curved surface, and the vertical plane face are maintained at the constant temperature $u = 0$ Over the upper base the temperature is a known function of position $f(r,\theta)$. Assuming steady-state conditions, find the temperature at any point in the solid

Because of the nature of the boundaries of the solid it will be highly inconvenient to use the heat equation in the cartesian form in which we derived it in Sec 8 2 Instead, we use it as expressed in cylindrical coordinates by means of the change of variables

$$x = r\cos\theta \qquad y = r\sin\theta \qquad z = z$$

namely,
$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\partial^2 u}{\partial z^2} = a^2\frac{\partial u}{\partial t}$$

or, more specifically, for steady-state conditions, under which $\dfrac{\partial u}{\partial t} = 0$,

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\partial^2 u}{\partial z^2} = 0 \tag{6}$$

Our first step is to assume a product solution $u(r,\theta,z) = R(r)\Theta(\theta)Z(z)$ and substitute it into (6) in an attempt to separate the variables This gives

$$R''\Theta Z + \frac{1}{r}R'\Theta Z + \frac{1}{r^2}R\Theta''Z + R\Theta Z'' = 0$$

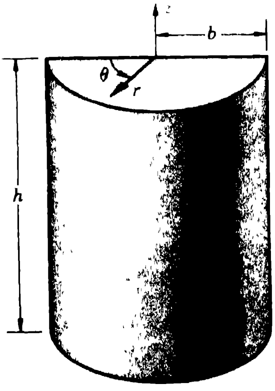


FIGURE 9.10
A half cylinder in which heat flow occurs because of surface temperature conditions.

or, multiplying by r^2 and dividing by $R\Theta Z$,

$$\frac{r^2R''}{R} + r\frac{R'}{R} + r^2\frac{Z''}{Z} = -\frac{\Theta''}{\Theta} = \mu_1$$

where the common value μ_1 is necessarily a constant, since the variables appearing on the respective sides of the equation are independent of each other.

If $\mu_1 < 0$, say $\mu_1 = -\nu^2$, then $\dfrac{\Theta''}{\Theta} = \nu^2$ and

$$\Theta = A\cosh\nu\theta + B\sinh\nu\theta \tag{7}$$

Now, by hypothesis,

$$u(r,0,z) = R(r)\Theta(0)Z(z) = 0 \qquad \text{and} \qquad u(r,\pi,z) = R(r)\Theta(\pi)Z(z) = 0$$

and these can hold for all values of r and z only if $\Theta(0) = \Theta(\pi) = 0$. From (7) we see that the condition $\Theta(0) = 0$ will be satisfied only if $A = 0$. To satisfy the condition $\Theta(\pi) = 0$, it is necessary that

$$B\sinh\nu\pi = 0$$

which, since $\nu \neq 0$, is possible only if $B = 0$. Thus, the possibility $\mu_1 < 0$ leads only to a trivial solution and, hence, must be rejected.

If $\mu_1 = 0$, then $\Theta'' = 0$ and

$$\Theta = A + B\theta$$

Again imposing the conditions $\Theta(0) = \Theta(\pi) = 0$, we find, as before, that $A = B = 0$. Hence, the possibility $\mu_1 = 0$ must also be rejected, since it leads only to a trivial solution.

Finally, if $\mu_1 > 0$, say $\mu_1 = \nu^2$, we have

$$\frac{\Theta''}{\Theta} = -\nu^2 \qquad \text{and} \qquad \Theta = A\cos\nu\theta + B\sin\nu\theta$$

For this to vanish when $\theta = 0$, we must have $A = 0$. For it to vanish when $\theta = \pi$, it is necessary that

$$B\sin\nu\pi = 0$$

Since we cannot permit B to be zero, because that would lead again to a trivial solution, we must have

$$\sin\nu\pi = 0$$

Hence $\qquad \nu = 1, 2, 3, \ldots$

and so for Θ we have the family of solutions

$$\Theta_n(\theta) = \sin n\theta$$

With μ_1 now known to be n^2, the differential equation for R and Z becomes

$$r^2\frac{R''}{R} + r\frac{R'}{R} + r^2\frac{Z''}{Z} = n^2$$

or, rearranging, $\qquad \dfrac{Z''}{Z} = \dfrac{n^2}{r^2} - \dfrac{R''}{R} - \dfrac{1}{r}\dfrac{R'}{R} = \mu_2$

where, again, since r and z are independent variables, it follows that the common value μ_2 must be a constant.

If $\mu_2 < 0$, say $\mu_2 = -\lambda^2$, we have

$$\frac{R''}{R} + \frac{1}{r}\cdot\frac{R'}{R} - \lambda^2 - \frac{n^2}{r^2} = 0 \qquad \text{or} \qquad r^2R'' + rR' - (\lambda^2r^2 + n^2)R = 0$$

which is precisely the modified Bessel equation. Hence,

$$R = CI_n(\lambda r) + DK_n(\lambda r)$$

Now $K_n(\lambda r)$ is infinite when $r = 0$, hence, to keep the temperature finite on the axis of the cylinder, it is necessary that $D = 0$ Also, by hypothesis,

$$u(b,\theta,z) = R(b)\Theta(\theta)Z(z) = 0$$

Hence, $$R(b) = CI_n(\lambda b) = 0$$

But the modified Bessel function I_n is never zero except possibly at the origin Therefore, the last condition can hold only if $C = 0$. But with C and D both zero, the solution is trivial, and so the possibility that $\mu_2 < 0$ must be rejected.

If $\mu_2 = 0$, then

$$\frac{R''}{R} + \frac{1}{r}\frac{R'}{R} - \frac{n^2}{r^2} = 0 \qquad \text{or} \qquad r^2R'' + rR' - n^2R = 0$$

This is not a Bessel-type equation, but is instead an example of the Euler equation (Example 3, Sec. 2.6) By the usual change of independent variable

$$r = e^v \qquad \text{or} \qquad v = \ln r$$

it becomes $$\frac{d^2R}{dv^2} - n^2R = 0$$

so that $$R = Ce^{nv} + De^{-nv} = Cr^n + Dr^{-n}$$

To keep the temperature finite on the axis, where $r = 0$, it is necessary that $D = 0$ To keep the temperature zero when $r = b$, it is necessary that

$$0 = Cb^n$$

which will be the case only if $C = 0$ This means that again the solution is trivial, and $\mu_2 = 0$ must also be rejected

Finally, if $\mu_2 > 0$, say $\mu_2 = \lambda^2$, we have

$$\frac{R''}{R} + \frac{1}{r}\frac{R'}{R} + \lambda^2 - \frac{n^2}{r^2} = 0 \qquad \text{or} \qquad r^2R'' + rR' + (\lambda^2r^2 - n^2)R = 0$$

and $$R = CJ_n(\lambda r) + DY_n(\lambda r)$$

Since $Y_n(\lambda r)$ is infinite when $r = 0$, we must have $D = 0$ To keep the temperature zero on the curved surface of the cylinder we must have

$$R(b) = CJ_n(\lambda b) = 0$$

Since $C = 0$ leads to a trivial solution, it is thus necessary that

$$J_n(\lambda b) = 0$$

that is, λ is restricted to the set of values

$$\left\{\frac{\rho_{nm}}{b}\right\}$$

where ρ_{nm} is the mth one of the roots of the equation $J_n(x) = 0$ Thus, for every value of n, there are infinitely many particular solutions for R, namely,

$$R_{nm}(r) = J_n(\lambda_{nm}r)$$

Now that we know that $\mu_2 = \lambda_{nm}^2$, it is an easy matter to solve for Z, and we have

$$\frac{Z''}{Z} = \lambda_{nm}^2 \qquad \text{and} \qquad Z = E\cosh\lambda_{nm}z + F\sinh\lambda_{nm}z$$

Since $u(r,\theta,0) \equiv R(r)\Theta(\theta)Z(0) = 0$, it follows that $Z(0) = 0$, from which we conclude that $E = 0$. The solution for Z associated with R_{nm} is, therefore,

$$Z_{nm}(z) = \sinh \lambda_{nm}z$$

For *each* n we therefore have infinitely many product solutions consisting of the same factor $\Theta(\theta) = \sin n\theta$ multiplied by the product of any pair of corresponding R's and Z's

$$u_{nm} = A_{nm}J_n(\lambda_{nm}r) \sinh \lambda_{nm}z \sin n\theta$$

In other words, we have a double array of product solutions,

$$\begin{array}{llll} u_{11}, & u_{12}, & u_{13}, & \ldots, u_{1m}, \ldots \\ u_{21}, & u_{22}, & u_{23}, & \ldots, u_{2m}, \ldots \\ \vdots \\ u_{n1}, & u_{n2}, & u_{n3}, & \ldots, u_{nm}, \ldots \\ \vdots \end{array}$$

Since none of the product solutions by itself is capable of representing the given temperature distribution $f(r,\theta)$ on the upper base, it is necessary that we construct an infinite series of the u_{nm}'s and try to make it fit the temperature condition when $z = h$. To build up a series for u we first add up all the product solutions associated with a particular value of n, getting

$$u_n = \sum_{m=1}^{\infty} u_{nm} = \sin n\theta \sum_{m=1}^{\infty} A_{nm}J_n(\lambda_{nm}r) \sinh \lambda_{nm}z$$

This, of course, amounts to forming the sums of the elements in each of the rows in the above array. Next, we add up all these series for every value of n

$$u(r,\theta,z) = \sum_{n=1}^{\infty} u_n = \sum_{n=1}^{\infty} \left[\sin n\theta \sum_{m=1}^{\infty} A_{nm}J_n(\lambda_{nm}r) \sinh \lambda_{nm}z\right] \tag{8}$$

The final step now is to determine the A's so that this double series will reduce to $f(r,\theta)$ when $z = h$

$$f(r,\theta) = \sum_{n=1}^{\infty} \left[\sin n\theta \sum_{m=1}^{\infty} A_{nm}J_n(\lambda_{nm}r) \sinh \lambda_{nm}h\right] \tag{9}$$

To carry out this expansion, let us imagine that r is held constant and that θ is allowed to vary over the range of the problem $(0,\pi)$. Under these conditions the inner sum in (9) is effectively a constant depending on n, say G_n, or more explicitly $G_n(r)$. That is,

$$f(r,\theta) = \sum_{n=1}^{\infty} G_n \sin n\theta$$

But the determination of the G's is a familiar problem! In fact it is nothing but the Fourier sine-expansion problem, and we can write immediately

$$G_n \equiv G_n(r) = \frac{2}{\pi}\int_0^{\pi} f(r,\theta) \sin n\theta \, d\theta \tag{10}$$

Thus $G_n(r)$ is a *known* function of r. But, by definition, $G_n(r)$ was the inner sum in (9), that is,

$$G_n(r) = \sum_{m=1}^{\infty} (A_{nm} \sinh \lambda_{nm}h)J_n(\lambda_{nm}r)$$

Hence, it is clear that the A's must be such that the products $A_{nm} \sinh \lambda_{nm}h$ are the coefficients in a Bessel function expansion of the now known function $G_n(r)$. Hence, from the theory of the last section, recalling that the λ's were determined by the condition

$$J_n(\lambda b) = 0$$

we can write

$$A_{nm} \sinh \lambda_{nm} h = \frac{\int_0^b r G_n(r) J_n(\lambda_{nm} r)\, dr}{(b^2/2) J_{n+1}^2(\lambda_{nm} b)}$$

Therefore,

$$A_{nm} = \frac{\int_0^b r G_n(r) J_n(\lambda_{nm} r)\, dr}{(b^2/2) \sinh \lambda_{nm} h\, J_{n+1}^2(\lambda_{nm} b)}$$

where $G_n(r)$ is given by (10). With the coefficients in the series solution (8) now determined, the problem is solved.

EXERCISES

1. What is $\mathcal{L}\{tJ_0(\lambda t)\}$?
2. What is $\mathcal{L}\{J_2(\lambda t)\}$? Hint: Recall from Eq. (3), Sec. 9.5, that
$$J_2(\lambda t) = J_0(\lambda t) - 2\frac{dJ_1(\lambda t)}{d(\lambda t)}$$
3. What is $\mathcal{L}\{J_n(\lambda t)\}$?
4. Show that $\int_0^\infty J_0(\lambda t)\, dt = \frac{1}{\lambda}$. [Hint: Consider the integral defining the Laplace transform of $J_0(\lambda t)$.]
5. What is (a) $\int_0^\infty J_1(\lambda t)\, dt$? (b) $\int_0^\infty tJ_0(\lambda t)\, dt$? (c) $\int_0^\infty tJ_1(\lambda t)\, dt$?
6. What is $\mathcal{L}^{-1}\left\{\frac{1}{\sqrt{s^2 + 4s + 13}}\right\}$?
7. What is $\mathcal{L}^{-1}\left\{\frac{1}{(s + a)\sqrt{s^2 + b^2}}\right\}$?
8. Show that $\mathcal{L}\{I_0(\lambda t)\} = \frac{1}{\sqrt{s^2 - \lambda^2}}$
9. What is (a) $\mathcal{L}\{tI_0(\lambda t)\}$? (b) $\mathcal{L}\{tI_1(\lambda t)\}$?
10. What is $\mathcal{L}\{I_1(\lambda t)\}$?
11. What is $\mathcal{L}^{-1}\left\{\frac{1}{\sqrt{s(s - 1)}}\right\}$?
12. Show that $\int_0^t J_0(\lambda) J_0(t - \lambda)\, d\lambda = \sin t$. (Hint: Recall the convolution theorem.)
13. Show that $I_0(t) = \frac{e^{-t}}{\pi}\int_0^t \frac{e^{2\lambda}}{\sqrt{\lambda(t - \lambda)}}\, d\lambda$. (Hint: Combine Formula 4, Sec. 7.3, for the case $n = -\frac{1}{2}$ with Theorem 5, Sec. 7.4, and then apply the convolution theorem to the result of Exercise 8.)
14. Find the solution of the equation $y'' + y = J_0(t)$ for which $y_0 = y_0' = 0$. Hence show that $tJ_1(t) = \int_0^t \sin(t - \lambda) J_0(\lambda)\, d\lambda$. [Hint: Solve the equation by Laplace transform methods using Eq. (1) and also the convolution theorem.]
15. What is $\int_0^t \sin(t - \lambda) J_1(\lambda)\, d\lambda$?
16. Derive Formula (1) by expressing $t^\nu J_\nu(\lambda t)$ as an infinite series and taking the Laplace transform term by term.
17. Show that $\mathcal{L}\{J_0(2\sqrt{t})\} = \frac{1}{s}e^{-1/s}$. What is $\mathcal{L}^{-1}\{e^{-1/s} - 1\}$?
18. Using Laplace transform methods, as illustrated in Example 1, show that a complete solution in terms of elementary functions can be obtained for the equation
$$(at + b)y'' + (ct + d)y' + (et + f)y = 0$$
if $ad - bc = 2a^2k$ and $af - be = ck$, where k is a negative integer.

19 Show that the function $\phi(z) = \int_0^\infty e^{-z\cosh\theta}\,d\theta$ satisfies the differential equation

$$z\phi'' + \phi' - z\phi = 0$$

Hence show that $\phi(z)$ is of the form $CK_0(z)$

20 In Example 3, verify that all the heat that enters the fin is lost from its surface What fraction of the heat entering the fin is lost from the section between $x = 0$ and $x = a/2$?

21 Work Example 3 given that the fin is of rectangular cross section

22 Show that the radial temperature distribution in a thin fin of rectangular cross section and outer radius R which completely encircles a heated cylinder of radius r satisfies the differential equation

$$\frac{d\left(x\dfrac{du}{dx}\right)}{dx} - \frac{2hx(u - u_0)}{kw} = 0$$

where x is measured radially outward from the center of the cylinder and the other parameters have the same significance as in Example 3

23 Solve the differential equation of Exercise 22, and find the temperature distribution in the fin if the cylinder temperature is u_c

24 Work Exercise 22, given that the fin is of triangular cross section

25 Find the first two natural frequencies of a steel shaft 20 in long vibrating torsionally if the shaft is built-in at one end and free at the other and if the radius of the shaft at a distance x from the free end is $r(x) = (x/20)^{1/4}$ Steel weighs 0 285 lb/in ^{3}, and its modulus of elasticity in shear is $E_s = 12 \times 10^6$ lb/in ^{2}

26 An elastic string whose weight per unit length is $w_0(1 + \alpha x)$, where x is the distance from one end of the string, is stretched under tension T between two points a distance l apart Find the equation defining the natural frequencies of the string

27 A body whose mass varies according to the law $m(t) = m_0(1 + \alpha t)^{-1}$ moves along the x-axis under the influence of a force of attraction which varies directly as the distance from the origin Determine the equation of motion of the body if it starts from rest at the point $x = x_0$

28 Work Exercise 27 if the force is directed away from the origin

29 The lower end of a long thin rod of uniform cross section is clamped so that the rod is vertical Determine the values of the parameters of the rod for which buckling will occur if the upper end of the rod is displaced slightly from its neutral position [Hint Choosing axes as in Example 1, Sec 2 6, the problem can be solved by using the relation $(EIy'')' = V$, where V is the transverse component of the weight of the portion of the rod above a general point x, or by using the relation $(EIy'')'' = -w$, where w is the transverse component of the weight per unit length of the rod at the point x]

30 A cantilever beam of length l and breadth b has its upper surface horizontal The depth of the beam varies directly as the cube root of the distance from the free end An oblique tensile force F, whose direction makes an angle θ with the horizontal, acts at the free end of the beam Find the equation of the deflection curve of the beam

31 Work Exercise 30 if the force is an oblique compressive force

32 A cantilever beam of length l and breadth b has its upper surface horizontal The depth of the beam varies directly as the two-thirds power of the distance from the free end If the beam bears a uniform load of w lb per unit length and is acted upon by a pure tensile force F at its free end, find the equation of the deflection curve of the beam

33 A bar has the shape of a truncated right circular cone of length l, the radii of its bases being r and R Find the frequency equation for the torsional vibrations of the bar, assuming both ends of the bar free

34 Determine the limiting form of the frequency equation in Exercise 33 when $r \to R$ Check by comparing your result with the frequency equation derived directly for a uniform bar (Hint: Express the Bessel functions in terms of sines and cosines)

35 Determine the natural frequencies of a uniform circular drumhead

36 Find the frequency equation for the transverse vibrations of a cantilever whose width is constant but whose depth varies directly as the distance from the free end (Hint. To solve the differential equation defining the normal modes of the beam, recall Exercise 8, Sec 9 4)

37 Find the frequency equation for the transverse vibrations of a cantilever beam which is a solid of revolution whose radius varies directly as the distance from the free end

38 Work Example 4 if the curved surface of the solid is perfectly insulated.

39 The lower base and curved surface of a right circular cylinder of radius b and height h are maintained at the constant temperature $u = 0$ Over the upper base the temperature is a known function of position $f(r,\theta)$ Assuming steady-state conditions, find the temperature at any point in the cylinder

40 A thin circular plate has its upper and lower faces insulated against the flow of heat One half of its circumference is maintained at the constant temperature 100°, the other half is maintained at the constant temperature 0° Find the steady-state temperature distribution in the plate

41 The region between two concentric circles of radii r_1 and r_2 is initially at a uniform temperature of zero At $t = 0$ the temperature around the entire inner boundary is suddenly raised to 100° Find the temperature at any point in the region at any subsequent time if the outer boundary is maintained at the temperature zero

42 A right circular cylinder of radius b and height h has its upper and lower bases maintained at the temperature 0° The curved surface of the cylinder is maintained at the temperature distribution $u(b,z) = f(z)$ Determine the steady-state temperature distribution throughout the cylinder

43 A right circular cylinder of radius b and height h has its lower base maintained at the constant temperature 0° Over its upper base the temperature distribution $u(r,h) = f(r)$ is maintained If the curved surface cools freely into air of constant temperature 0°, find the steady-state temperature distribution within the cylinder

44 A two-dimensional region having the shape of a quarter of a circle is initially at a uniform temperature of 100° At $t = 0$ the temperature around the entire boundary is suddenly reduced to zero and maintained thereafter at that value Find the temperature at any point of the region at any subsequent time

45 Find the steady-state temperature distribution in a two-dimensional region having the shape of a quarter of a circle if the curved boundary and one of the radial boundaries is maintained at the constant temperature 0° and the other radial boundary is maintained at the constant temperature 100°.

9.8

Legendre polynomials

In Example 4, Sec 9.7, in solving the steady-state heat equation, i e , Laplace's equation, in cylindrical coordinates, we found that one of the ordinary differential equations arising from the separation of variables was Bessel's equation. In very much the same way, it turns out that, when we apply the method of separation of variables to Laplace's equation in spherical coordinates, one of the ordinary differential equations which results is *Legendre's equation.*

If the expression

$$\nabla^2 F = \frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + \frac{\partial^2 F}{\partial z^2}$$

is transformed from cartesian coordinates to spherical coordinates by means of the relations (Fig. 9.11)

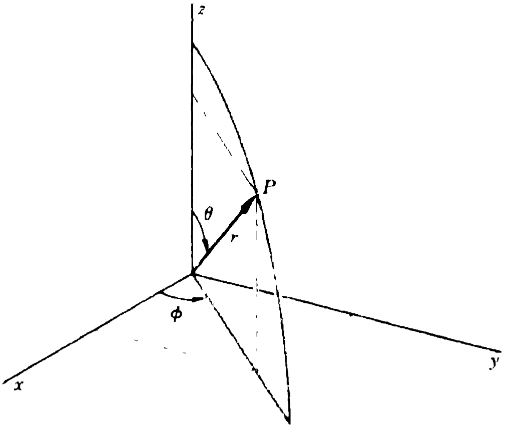


FIGURE 9.11 Plot showing the relation between rectangular and spherical coordinates

$$x = r \sin\theta \cos\phi \qquad y = r \sin\theta \sin\phi \qquad z = r\cos\theta$$

we obtain, after a lengthy but straightforward reduction,

$$\nabla^2 F = \frac{1}{r^2 \sin\theta}\left(r^2 \sin\theta \frac{\partial^2 F}{\partial r^2} + 2r\sin\theta\frac{\partial F}{\partial r} + \sin\theta\frac{\partial^2 F}{\partial\theta^2} + \cos\theta\frac{\partial F}{\partial\theta} + \frac{1}{\sin\theta}\frac{\partial^2 F}{\partial\phi^2}\right)$$

Hence, when Laplace's equation $\nabla^2 F = 0$ is expressed in spherical coordinates, it becomes

$$r^2 \sin\theta \frac{\partial^2 F}{\partial r^2} + 2r\sin\theta\frac{\partial F}{\partial r} + \sin\theta\frac{\partial^2 F}{\partial\theta^2} + \cos\theta\frac{\partial F}{\partial\theta} + \frac{1}{\sin\theta}\frac{\partial^2 F}{\partial\phi^2} = 0 \tag{1}$$

Any solution $F(r,\theta,\phi)$ of this equation is known as a **spherical harmonic.**

In an attempt to solve Eq. (1), we assume a product solution

$$F(r,\theta,\phi) = R(r)G(\theta,\phi)$$

Then, substituting this into (1), we have

$$r^2 \sin\theta\, R''G + 2r\sin\theta\, R'G + \sin\theta\, R\frac{\partial^2 G}{\partial\theta^2} + \cos\theta\, R\frac{\partial G}{\partial\theta} + \frac{R}{\sin\theta}\frac{\partial^2 G}{\partial\phi^2} = 0$$

or, dividing through by $RG\sin\theta$ and rearranging,

$$\frac{r^2R'' + 2rR'}{R} = -\left(\frac{1}{G}\frac{\partial^2 G}{\partial\theta^2} + \frac{\cos\theta}{G\sin\theta}\frac{\partial G}{\partial\theta} + \frac{1}{G\sin^2\theta}\frac{\partial^2 G}{\partial\phi^2}\right)$$

This relation can hold only if the common value of these two expressions is a constant. For later convenience we write the constant as $n(n+1)$; hence, we are led to the two equations

$$r^2R'' + 2rR' - n(n+1)R = 0 \tag{2}$$

$$\frac{\partial^2 G}{\partial\theta^2} + \frac{\cos\theta}{\sin\theta}\frac{\partial G}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2 G}{\partial\phi^2} + n(n+1)G = 0 \tag{3}$$

The first of these equations is an instance of Euler's equation (Example 3, Sec 2 6), and it is easy to verify that its complete solution is

$$R = c_1 r^n + \frac{c_2}{r^{n+1}}$$

Solutions $G(\theta,\phi)$ of the second equation, which we will have to find by a further separation of variables, are known as **surface harmonics.**

If, in (3), we substitute $G(\theta,\phi) = \Theta(\theta)\Phi(\phi)$, we find

$$\Theta''\Phi + \frac{\cos\theta}{\sin\theta}\Theta'\Phi + \frac{1}{\sin^2\theta}\Theta\Phi'' + n(n+1)\Theta\Phi = 0$$

or, dividing by $\Theta\Phi/\sin^2\theta$ and rearranging slightly,

$$\sin^2\theta\frac{\Theta''}{\Theta} + \sin\theta\cos\theta\frac{\Theta'}{\Theta} + n(n+1)\sin^2\theta = -\frac{\Phi''}{\Phi}$$

Again, the common value of the two members of this equation must be a constant, say m^2, thus, we have the pair of equations

(4) $$\Phi'' + m^2\Phi = 0$$

(5) $$\sin^2\theta\,\Theta'' + \sin\theta\cos\theta\,\Theta' + [n(n+1)\sin^2\theta - m^2]\Theta = 0$$

The first of these equations is completely familiar; and its complete solution

$$\Phi = c_3\cos m\phi + c_4\sin m\phi$$

can be written down at once The second equation is known as the **associated Legendre equation,*** although it is usually studied in the form obtained by setting $x = \cos\theta$

If $x = \cos\theta$, then

$$\frac{d\Theta}{d\theta} = \frac{d\Theta}{dx}\frac{dx}{d\theta} = -\sin\theta\frac{d\Theta}{dx}$$

$$\frac{d^2\Theta}{d\theta^2} = \frac{d}{d\theta}\left(-\sin\theta\frac{d\Theta}{dx}\right) = -\cos\theta\frac{d\Theta}{dx} - \sin\theta\frac{d^2\Theta}{dx^2}\frac{dx}{d\theta}$$

$$= -\cos\theta\frac{d\Theta}{dx} + \sin^2\theta\frac{d^2\Theta}{dx^2}$$

Hence, substituting these expressions into (5), we obtain the equation

$$\sin^2\theta\left(-\cos\theta\frac{d\Theta}{dx} + \sin^2\theta\frac{d^2\Theta}{dx^2}\right) + \sin\theta\cos\theta\left(-\sin\theta\frac{d\Theta}{dx}\right) + [n(n+1)\sin^2\theta - m^2]\Theta = 0$$

or, dividing out $\sin^2\theta$, substituting $x = \cos\theta$ in the coefficients, and simplifying,

(6) $$(1-x^2)\frac{d^2\Theta}{dx^2} - 2x\frac{d\Theta}{dx} + \left[n(n+1) - \frac{m^2}{1-x^2}\right]\Theta = 0$$

---

* After the French mathematician Adrien-Marie Legendre (1752–1833).

This is the **algebraic form of the associated Legendre equation.** If $m = 0$, that is, if the solution of the original problem is independent of the longitude angle ϕ, then Eq. (6) reduces to

(7) $$(1 - x^2)\frac{d^2\Theta}{dx^2} - 2x\frac{d\Theta}{dx} + n(n+1)\Theta = 0$$

which is known simply as **Legendre's equation.**

To solve Eq. (7) we use the method of Frobenius and assume a series solution of the form

$$\Theta(x) = x^c(a_0 + a_1x + a_2x^2 + \cdots + a_kx^k + \cdots) \qquad a_0 \neq 0$$

Then substituting into Eq. (7), we have

$$\begin{array}{rr}
a_0c(c-1)x^{c-2} + a_1(c+1)cx^{c-1} + a_2(c+2)(c+1)x^c + \cdots & + a_{k+2}(c+k+2)(c+k+1)x^{c+k} + \cdots \\
- a_0c(c-1)x^c - \cdots & - a_k(c+k)(c+k-1)x^{c+k} - \cdots \\
- 2a_0cx^c - \cdots & - 2a_k(c+k)x^{c+k} - \cdots \\
+ n(n+1)a_0x^c + \cdots & + n(n+1)a_kx^{c+k} + \cdots
\end{array} = 0$$

For this to be an identity, it is necessary that

$$a_0c(c-1) = 0 \qquad a_1(c+1)c = 0$$

and, in general,

$$a_{k+2}(c+k+2)(c+k+1) - a_k[(c+k)(c+k+1) - n(n+1)] = 0$$

If we take $c = 0$, both a_0 and a_1 remain arbitrary, and we have for the general recurrence relation

(8) $$a_{k+2} = -\frac{(n-k)(n+k+1)}{(k+1)(k+2)}a_k \qquad k = 0, 1, 2, \ldots$$

Specifically, from (8),

$$a_0 = a_0 \qquad\qquad a_1 = a_1$$

$$a_2 = -\frac{n(n+1)}{2!}a_0 \qquad\qquad a_3 = -\frac{(n-1)(n+2)}{3!}a_1$$

$$a_4 = \frac{n(n-2)(n+1)(n+3)}{4!}a_0 \qquad\qquad a_5 = \frac{(n-1)(n-3)(n+2)(n+4)}{5!}a_1$$

$$\cdots \qquad\qquad \cdots$$

Hence a complete solution of (7) can be written

(9) $$\Theta(x) = a_0\left[1 - \frac{n(n+1)}{2!}x^2 + \frac{n(n-2)(n+1)(n+3)}{4!}x^4 - \cdots\right]$$
$$+ a_1\left[x - \frac{(n-1)(n+2)}{3!}x^3 + \frac{(n-1)(n-3)(n+2)(n+4)}{5!}x^5 - \cdots\right]$$

These infinite series define what are known as **Legendre functions of the second kind.** Since $x = \pm 1$ are the only singular points of the differential equation (7), it follows from Theorem 1, Sec. 9.1, that the radius of convergence of these series is 1. It can be shown, however, that neither series converges at either of the end points $x = \pm 1$; that is, the interval of convergence for each series is $-1 < x < 1$.

In many applications the parameter n is a positive integer If it is odd, then, clearly, the second series in (9) contains only a finite number of terms, if it is even, then the first series contains only a finite number of terms In either of these cases the series which reduces to a finite sum is known as a **Legendre polynomial** or **zonal harmonic of order** n. To obtain a standard form for the Legendre polynomials it is customary to multiply the finite sums occurring in (9) when n is an integer by the appropriate one of the following factors

$$(-1)^{n/2}\frac{1\cdot 3\cdot 5\cdots(n-1)}{2\cdot 4\cdot 6\cdots n} = \frac{(-1)^{n/2}n!}{2^n\left[\left(\frac{n}{2}\right)!\right]^2} \qquad n \text{ even}$$

$$(-1)^{(n-1)/2}\frac{1\cdot 3\cdot 5\cdots n}{2\cdot 4\cdot 6\cdots(n-1)} = \frac{(-1)^{(n-1)/2}(n+1)!}{2^n\left(\frac{n-1}{2}\right)!\left(\frac{n+1}{2}\right)!} \qquad n \text{ odd}$$

This leads to the general formula

$$\text{(10)} \qquad P_n(x) = \sum_{k=0}^{N}\frac{(-1)^k(2n-2k)!}{2^n k!(n-k)!(n-2k)!}x^{n-2k} \qquad \begin{cases} N = \dfrac{n}{2} & n \text{ even} \\ N = \dfrac{n-1}{2} & n \text{ odd} \end{cases}$$

Specifically, we have

$$P_0(x) = 1 \qquad P_1(x) = x$$

$$P_2(x) = \tfrac{1}{2}(3x^2 - 1) \qquad P_3(x) = \tfrac{1}{2}(5x^3 - 3x)$$

$$P_4(x) = \tfrac{1}{8}(35x^4 - 30x^2 + 3) \qquad P_5(x) = \tfrac{1}{8}(63x^5 - 70x^3 + 15x)$$

As these particular results illustrate, $P_n(1) = 1$ and $P_n(-1) = (-1)^n$ for all values of n Since the infinite series in (9) diverge when $x = \pm 1$, it is clear that *to within an arbitrary constant multiplier, $P_n(x)$ is the only solution of Legendre's equation which is finite on the closed interval $-1 \leqq x \leqq 1$*

One of the fundamental identities involving Legendre polynomials is **Rodrigues' formula**:*

THEOREM 1

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n(x^2-1)^n}{dx^n}$$

PROOF This result can be proved by direct differentiation and induction, but it is perhaps more interesting to proceed in the following way If we let

$$v = (x^2 - 1)^n$$

then

$$\frac{dv}{dx} = 2nx(x^2 - 1)^{n-1}$$

---

* Named for the French economist and mathematician Olinde Rodrigues (1794–1851).

or, multiplying the last equation by $x^2 - 1$,

$$(x^2 - 1)\frac{dv}{dx} = 2nx(x^2 - 1)^n$$

and, finally, $$(1 - x^2)\frac{dv}{dx} + 2nxv = 0$$

If we differentiate this repeatedly with respect to x, we obtain

$$(1 - x^2)v'' + 2(n - 1)xv' + (2n)v = 0$$
$$(1 - x^2)v''' + 2(n - 2)xv'' + 2(2n - 1)v' = 0$$
$$\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot$$

and, after $k + 1$ differentiations,

$$(1 - x^2)v^{(k+2)} + 2(n - k - 1)xv^{(k+1)} + (k + 1)(2n - k)v^{(k)} = 0$$

If we now take $k = n$ and put $v^{(k)} = u$, the last equation becomes

$$(1 - x^2)u'' - 2xu' + n(n + 1)u = 0$$

which is precisely Legendre's equation But

$$u = v^{(n)} = \frac{d^n(1 - x^2)^n}{dx^n}$$

is obviously a polynomial of degree n. Moreover, from (9) it is clear that to within a constant factor there is only one polynomial solution of Legendre's equation, namely, $P_n(x)$ Hence, $P_n(x)$ must be some multiple of u, that is,

$$P_n(x) = c\frac{d^n(1 - x^2)^n}{dx^n}$$

Finally, we can determine c by equating the coefficients of x^n in the two members of the last identity Clearly, the coefficient of x^n on the right-hand side is

$$(-1)^n c(2n)(2n - 1) \cdots [2n - (n - 1)] = (-1)^n \frac{(2n)!}{n!} c$$

Moreover, setting $k = 0$ in Eq (10), we see that the coefficient of x^n in $P_n(x)$ is

$$\frac{(2n)!}{2^n(n!)^2}$$

Hence, we must have

$$\frac{(2n)!}{2^n(n!)^2} = (-1)^n\frac{(2n)!}{n!}c \qquad \text{or} \qquad c = \frac{(-1)^n}{2^n n!}$$

which completes the verification of Rodrigues' formula

Another important identity involving Legendre polynomials is embodied in the following theorem

THEOREM 2

$$\frac{1}{\sqrt{1 - 2xz + z^2}} = P_0(x) + P_1(x)z + P_2(x)z^2 + \cdots + P_n(x)z^n + \cdots$$

PROOF To prove this, we expand the radical on the left-hand side by the binomial theorem, getting

$$
\begin{aligned}
[1 - z(2x - z)]^{-1/2} = 1 &+ \frac{1}{2} z(2x - z) + \frac{1 \cdot 3}{2^2 2!} z^2(2x - z)^2 + \cdots \\
&+ \frac{1 \cdot 3 \cdots (2n-3)}{2^{n-1}(n-1)!} z^{n-1}(2x - z)^{n-1} \\
&+ \frac{1 \cdot 3 \cdots (2n-1)}{2^n n!} z^n(2x - z)^n + \cdots
\end{aligned}
$$

Now z^n can occur only in the terms out to and including the one containing $z^n(2x - z)^n$, and from these, by expanding the various powers of $(2x - z)$, we find that its total coefficient is

$$
\begin{aligned}
&\frac{1 \cdot 3 \cdots (2n-1)}{2^n n!} (2x)^n - \frac{1 \cdot 3 \cdots (2n-3)}{2^{n-1}(n-1)!} \frac{n-1}{1!} (2x)^{n-2} \\
&\qquad + \frac{1 \cdot 3 \cdots (2n-5)}{2^{n-2}(n-2)!} \frac{(n-2)(n-3)}{2!} (2x)^{n-4} - \cdots
\end{aligned}
$$

or, multiplying and dividing by the factors needed to complete the factorials in the numerators,

$$
\frac{(2n)!}{2^n n! n!} x^n - \frac{(2n-2)!}{2^{n-1} 1! (n-1)! (n-2)!} x^{n-2} + \frac{(2n-4)!}{2^{n-2} 2! (n-2)! (n-4)!} x^{n-4} - \cdots
$$

which is precisely the expanded form of $P_n(x)$, as given by (10) Thus

$$
(1 - 2xz + z^2)^{-1/2} = \sum_{n=0}^{\infty} P_n(x) z^n \qquad \text{as asserted.}
$$

In other words, the expression $(1 - 2xz + z^2)^{-1/2}$ is a *generating function* for the Legendre polynomials, analogous to the generating function $\exp \dfrac{x}{2}\left(t - \dfrac{1}{t}\right)$ for the Bessel functions which we investigated in Sec. 9.5

In many applications the algebraic form of the Legendre polynomials is the more useful. There are problems, however, in which it is essential that they be expressed in terms of θ, the colatitude angle of the spherical coordinate system with which our discussion began. This can easily be done by reversing the transformation $x = \cos\theta$ which led from the trigonometric to the algebraic form of Legendre's equation However, replacing x by $\cos\theta$ in $P_n(x)$ leads to expressions which are quite inconvenient because of the powers of $\cos\theta$ they contain Fortunately, using the generating function provided by Theorem 2, we can easily derive more useful forms in which cosines of multiples of θ take the place of powers of $\cos\theta$.

To do this, let us substitute

$$
x = \cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}
$$

into the generating function, getting

$$
[1 - z(e^{i\theta} + e^{-i\theta}) + z^2]^{-1/2} = [(1 - ze^{i\theta})(1 - ze^{-i\theta})]^{-1/2} = \sum_{n=1}^{\infty} P_n z^n
$$

Now, if we use the binomial theorem to expand each of the factors in the middle term of this continued identity, we obtain

$$(1 - ze^{i\theta})^{-1/2} = 1 + \frac{1}{2} ze^{i\theta} + \frac{1 \cdot 3}{2 \cdot 4} z^2 e^{2i\theta} + \cdots + \frac{1 \cdot 3 \cdots (2n-1)}{2 \cdot 4 \cdots (2n)} z^n e^{ni\theta} + \cdots$$

and

$$(1 - ze^{-i\theta})^{-1/2} = 1 + \frac{1}{2} ze^{-i\theta} + \frac{1 \cdot 3}{2 \cdot 4} z^2 e^{-2i\theta} + \cdots + \frac{1 \cdot 3 \cdots (2n-1)}{2 \cdot 4 \cdots (2n)} z^n e^{-ni\theta} + \cdots$$

The coefficient of z^n in the product of these two series is easy to determine, and we find for it the expression

$$\frac{1 \cdot 3 \cdots (2n-1)}{2 \cdot 4 \cdots (2n)} (e^{ni\theta} + e^{-ni\theta}) + \frac{1}{2} \frac{1 \cdot 3 \cdots (2n-3)}{2 \cdot 4 \cdots (2n-2)} (e^{(n-2)i\theta} + e^{-(n-2)i\theta})$$
$$+ \frac{1 \cdot 3}{2 \cdot 4} \frac{1 \cdot 3 \cdots (2n-5)}{2 \cdot 4 \cdots (2n-4)} (e^{(n-4)i\theta} + e^{-(n-4)i\theta}) + \cdots$$

Hence, replacing the various combinations of exponentials by their cosine equivalents and recalling that the coefficient of z^n in the expansion of the generating function is just P_n, we have finally

$$P_n(\cos\theta) = \frac{1 \cdot 3 \cdots (2n-1)}{2 \cdot 4 \cdots (2n)} 2 \cos n\theta \tag{11}$$
$$+ \frac{1}{2} \frac{1 \cdot 3 \cdots (2n-3)}{2 \cdot 4 \cdots (2n-2)} 2 \cos (n-2)\theta$$
$$+ \frac{1 \cdot 3}{2 \cdot 4} \frac{1 \cdot 3 \cdots (2n-5)}{2 \cdot 4 \cdots (2n-4)} 2 \cos (n-4)\theta + \cdots$$

If n is odd, the final term in $P_n(\cos\theta)$ contains the factor $\cos\theta$ and is correctly given by the last nonzero term in the series (11). However, if n is even, the final term in $P_n(\cos\theta)$ is a constant which is equal to just half the last nonzero term in the series (11). This is the case because, although the general term in the coefficient of z^n contains both $e^{(n-2k)i\theta}$ and $e^{-(n-2k)i\theta}$, when n is even and $k = n/2$ these terms are identical and arise only once in the product of the series for $(1 - ze^{i\theta})^{-1/2}$ and $(1 - ze^{-i\theta})^{-1/2}$ and not twice. Thus, specifically,

$$P_0(\cos\theta) = 1$$

$$P_1(\cos\theta) = \cos\theta$$

$$P_2(\cos\theta) = \frac{3\cos 2\theta + 1}{4}$$

$$P_3(\cos\theta) = \frac{5\cos 3\theta + 3\cos\theta}{8}$$

$$P_4(\cos\theta) = \frac{35\cos 4\theta + 20\cos 2\theta + 9}{64}$$

$$P_5(\cos\theta) = \frac{63\cos 5\theta + 35\cos 3\theta + 30\cos\theta}{128}$$

. . . .

Since Legendre's equation can be written in the form

$$\frac{d[(1-x^2)y']}{dx} + n(n+1)y = 0$$

it is clear that it is a special case, with

$$p(x) = 1 \qquad q(x) = 0 \qquad r(x) = 1 - x^2 \qquad \lambda = n(n+1)$$

of the equation covered by Theorem 4, Sec 8 5. Hence, if solutions of Legendre's equation satisfy suitable boundary conditions, they must be orthogonal In particular, for the important interval $(-1,1)$ no boundary conditions are necessary, since $r(x) \equiv 1 - x^2$ vanishes at each end point; that is,

$$\int_{-1}^{1} P_m(x)P_n(x)\,dx = 0 \qquad m \neq n \tag{12}$$

Before the property of orthogonality can be used to expand an arbitrary function in terms of Legendre polynomials, we must, of course, know the value of the integral of the square of the general Legendre polynomial. This can be obtained in various ways, but perhaps the simplest is to use the generating function provided by Theorem 2 If we square the identity

$$\frac{1}{(1-2xz+z^2)^{1/2}} = P_0(x) + P_1(x)z + \cdots + P_n(x)z^n + \cdots$$

and integrate with respect to x from -1 to 1, we obtain

$$\int_{-1}^{1} \frac{dx}{1-2xz+z^2} = \int_{-1}^{1} [P_0(x) + P_1(x)z + \cdots + P_n(x)z^n + \cdots]^2\,dx$$

The integral on the left is easily evaluated On the right, all integrals involving the product of two different P's are zero because of the orthogonality property (12) Hence,

$$-\frac{1}{2z}\ln(1-2xz+z^2)\Big|_{-1}^{1} = \int_{-1}^{1} P_0^2(x)\,dx + z^2\int_{-1}^{1} P_1^2(x)\,dx + \cdots + z^{2n}\int_{-1}^{1} P_n^2(x)\,dx + \cdots \tag{13}$$

Evaluation of the left member leads at once to

$$-\frac{1}{2z}[\ln(1-z)^2 - \ln(1+z)^2] = \frac{1}{z}[\ln(1+z) - \ln(1-z)]$$

Moreover, if we replace the logarithms by their respective power series, we obtain

$$\frac{1}{z}\left(z - \frac{z^2}{2} + \frac{z^3}{3} - \cdots - \frac{z^{2n}}{2n} + \frac{z^{2n+1}}{2n+1} - \cdots\right)$$
$$-\frac{1}{z}\left(-z - \frac{z^2}{2} - \frac{z^3}{3} - \cdots - \frac{z^{2n}}{2n} - \frac{z^{2n+1}}{2n+1} - \cdots\right)$$
$$= 2\left(1 + \frac{z^2}{3} + \frac{z^4}{5} + \cdots + \frac{z^{2n}}{2n+1} + \cdots\right)$$

Hence, comparing coefficients of z^{2n} in this series and in the right member of (13), we obtain the desired result:

$$\int_{-1}^{1} P_n^2(x)\,dx = \frac{2}{2n+1} \tag{14}$$

By means of the substitution $x = \cos\theta$, Eqs. (12) and (13) can be transformed at once into corresponding results for the Legendre polynomials in trigonometric form. Hence, we can state the following important theorem:

THEOREM 3

The Legendre polynomials in algebraic form satisfy the orthogonality relations

$$\int_{-1}^{1} P_m(x)P_n(x)\,dx = \begin{cases} 0 & m \neq n \\ \dfrac{2}{2n+1} & m = n \end{cases}$$

In trigonometric form, the Legendre polynomials satisfy the orthogonality relations

$$\int_0^{\pi} P_m(\cos\theta)P_n(\cos\theta)\sin\theta\,d\theta = \begin{cases} 0 & m \neq n \\ \dfrac{2}{2n+1} & m = n \end{cases}$$

EXAMPLE 1

The known temperature distribution $u = f(\theta)$ is maintained over the entire surface of a sphere of radius b. Find the steady-state temperature at any point in the sphere.

Here we have to solve the steady-state heat equation, i.e., Laplace's equation, in spherical coordinates. However, from the obvious circular symmetry of the problem it is clear that u is a function of r and θ only. Hence $\dfrac{\partial^2 u}{\partial\phi^2} = 0$, and Eq. (1) reduces to

$$r^2\sin\theta\frac{\partial^2 u}{\partial r^2} + 2r\sin\theta\frac{\partial u}{\partial r} + \sin\theta\frac{\partial^2 u}{\partial\theta^2} + \cos\theta\frac{\partial u}{\partial\theta} = 0 \tag{15}$$

Assuming a product solution

$$u = R(r)\Theta(\theta)$$

and substituting into (15), we obtain

$$r^2\sin\theta\,R''\Theta + 2r\sin\theta\,R'\Theta + \sin\theta\,R\Theta'' + \cos\theta\,R\Theta' = 0$$

From this, by dividing by $\sin\theta\,R\Theta$ and transposing, we have

$$\frac{r^2R''}{R} + \frac{2rR'}{R} = -\frac{\Theta''}{\Theta} - \frac{\cos\theta}{\sin\theta}\frac{\Theta'}{\Theta} = \nu$$

Since for any ν, the quadratic equation $n^2 + n - \nu = 0$ is always satisfied by at least one (possibly complex) value of n, it is no specialization to take $\nu = n(n+1)$, so that we have the two ordinary differential equations

$$r^2R'' + 2rR' - n(n+1)R = 0$$

$$\sin\theta\,\Theta'' + \cos\theta\,\Theta' + n(n+1)\sin\theta\,\Theta = 0$$

The first of these is just an instance of Euler's equation, and its general solution is easily found to be

$$R = Ar^n + \frac{B}{r^{n+1}}$$

However, since we require solutions which are finite when $r = 0$, it is clear that we must specialize this by taking $B = 0$ The second equation is Legendre's equation. Since we require solutions of it which are finite over the closed interval $0 \leq \theta \leq \pi$ and since the only such solutions are the Legendre polynomials $P_n(\cos\theta)$, it is clear that n must be an integer and

$$\Theta = P_n(\cos\theta)$$

Hence, we have the infinite sequence of product solutions

$$A_1 r P_1(\cos\theta) \qquad A_2 r^2 P_2(\cos\theta) \qquad . \qquad A_n r^n P_n(\cos\theta)$$

None of these by itself can satisfy the given temperature condition

$$u(b,\theta) = f(\theta)$$

on the surface of the sphere Hence, as usual, we form an infinite series of the individual product solutions and attempt to make it fit the boundary condition Thus we write

$$u(r,\theta) = \sum_{n=1}^{\infty} A_n r^n P_n(\cos\theta) \tag{16}$$

Then, substituting $r = b$ and $u(b,\theta) = f(\theta)$, we get

$$f(\theta) = \sum_{n=1}^{\infty} A_n b^n P_n(\cos\theta)$$

To find A_n we multiply the last equation by $\sin\theta\, P_n(\cos\theta)$ and integrate from 0 to π By virtue of the orthogonality properties of the P's, all integrals on the right except one become zero, and we have

$$\int_0^\pi f(\theta)\sin\theta\, P_n(\cos\theta)\,d\theta = A_n b^n \frac{2}{2n+1}$$

or

$$A_n = \frac{2n+1}{2b^n}\int_0^\pi f(\theta)\sin\theta\, P_n(\cos\theta)\,d\theta$$

With the coefficients in the series (16) known, the problem is now formally solved

EXERCISES

1 a Show that any polynomial in x of degree m can be represented uniquely by a finite sum of the form

$$a_0P_0(x) + a_1P_1(x) + \qquad + a_mP_m(x)$$

Hence show that

$$\int_{-1}^{1} x^m P_n(x)\,dx = 0 \qquad \text{if } m < n$$

b Express x^3 and x^4 as linear combinations of Legendre polynomials

2 Consider the functions x^i $(i = 0, 1, 2, 3, \qquad)$, and let $f_n(x) = \sum_{i=0}^{n} a_{ni}x^i$, where the a's are coefficients to be determined so that

$$\int_{-1}^{1} f_m(x)f_n(x)\,dx = \begin{cases} 0 & m \neq n \\ \dfrac{2}{2n+1} & m = n \end{cases}$$

Calculate the a's for $n = 0, 1, 2, 3$, and show that, at least for these values of n, $f_n(x) = P_n(x)$

3 It is desired to approximate a function $f(x)$ over the interval $(-1,1)$ by a polynomial $P(x)$ of degree n which will make the integral

$$\int_{-1}^{1} [f(x) - P(x)]^2\,dx$$

a minimum Show that $P(x)$ is the nth partial sum of the expansion of $f(x)$ over the interval $(-1,1)$ in terms of Legendre polynomials. (The Legendre polynomials thus play the same role in the least-square approximation of continuous functions that the orthogonal polynomials discussed in Sec 4 6 play in the least-square approximation of tabular functions)

4 Show that the Legendre polynomials with even subscripts and the Legendre polynomials with odd subscripts both form orthogonal sets over the interval $(0,1)$

5 By differentiating the generating function for the Legendre polynomials, show that all derivatives of even order of $P_n(x)$ vanish at $x = 0$ if n is odd, and that all derivatives of odd order vanish at $x = 0$ if n is even What are the values of the nonzero derivatives at $x = 0$?

6 Using Rodrigues' formula, prove that

$$P'_{n+1}(x) - P'_{n-1}(x) = (2n + 1)P_n(x)$$

Hence show that

$$\int_x^1 P_n(x)\,dx = \frac{P_{n-1}(x) - P_{n+1}(x)}{2n+1}$$

7 Find the steady-state temperature at any point in a spherical shell of inner radius b_1 and outer radius b_2 if the temperature distributions $u(b_1,\theta) = f_1(\theta)$ and $u(b_2,\theta) = f_2(\theta)$ are maintained over the inner and outer surfaces, respectively.

8 The temperature distribution $u(b,\theta) = f(\theta)$ is maintained over the curved surface of a hemisphere of radius b The plane boundary of the hemisphere is kept at the temperature $u = 0$ Find the steady-state temperature at any point in the hemisphere. (Hint: The results of Exercise 4 may be of assistance)

9 Find polynomial solutions of the equation

$$y'' - 2xy + 2ny = 0 \qquad n \text{ an integer}$$

and show that they are orthogonal with respect to the weight function e^{-x^2} over the interval $(-\infty,\infty)$ [This equation is known as **Hermite's equation**, after the French mathematician Charles Hermite (1822–1901), and its polynomial solutions are known as **Hermite polynomials.**]

10 Find polynomial solutions of the equation

$$xy'' + (1 - x)y' + ny = 0 \qquad n \text{ an integer}$$

and show that they are orthogonal with respect to the weight function e^{-x} over the interval $(0,\infty)$ [This equation is known as **Laguerre's equation**, after the French mathematician Edmond Laguerre (1834–1886), and its polynomial solutions are known as **Laguerre polynomials.**]

CHAPTER TEN

Determinants and Matrices

10.1

Determinants

In a restricted sense, at least, the concept of a determinant is already familiar from elementary algebra, where, in solving systems of two and three simultaneous linear equations, we found it convenient to introduce what we called *determinants of the second and third order.* In the work of this book we shall have occasion to generalize these ideas to the solution of systems of more than three linear equations and to other applications not immediately associated with solving equations For this reason we shall devote this and the following chapter to a review and an extension of our earlier study of determinants and to a discussion of some of the fundamental properties of the related mathematical objects known as *matrices*

By a **determinant of order** n we mean a certain function of n^2 quantities, which we shall describe more precisely as soon as we have introduced the necessary notation and preliminary definitions. The customary symbol for a determinant consists of a square array of the n^2 quantities enclosed between vertical bars

$$(1)\qquad |A|\dagger = |a_{ij}| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdot & a_{2n} \\ \cdot\;\cdot & \cdot\;\cdot & \cdot\;\cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

For brevity, we shall often use the word *determinant* to refer to this symbol as well as to the expansion‡ for which it stands.

---

† The use of vertical bars in the notation for a determinant and in the notation for the absolute value of a quantity, while perhaps unfortunate, is universal. Which meaning is intended in any particular case should always be clear from the context.

‡ See Definition 1, this section, p. 403.

Although logically undesirable, this dual usage is quite common and should cause no confusion

The quantities a_{ij} which appear in (1) are called the **elements** of the determinant The horizontal lines of elements are called **rows**; the vertical lines of elements are called **columns.** In the convenient **double subscript notation** illustrated in (1), the first subscript associated with an element identifies the row and the second subscript identifies the column in which the element lies. There is, of course, no reason to suppose that the element in the ith row and jth column is the same as the element in the jth row and ith column, and so in general $a_{ij} \neq a_{ji}$ The sloping line of elements extending from a_{11} to a_{nn} is called the **principal diagonal** of the determinant.

The determinant $|M|$ formed by the m^2 elements common to any m rows and any m columns of an nth-order determinant $|A|$ is said to be an **mth-order minor** of $|A|$. The determinant of order $n - m$ formed by the elements which remain when the m rows and m columns containing an mth-order minor $|M|$ are deleted from $|A|$ is called the **complementary minor** of $|M|$. If the numbers of the rows and columns of $|A|$ which contain an mth-order minor $|M|$ are, respectively,

$$i_1, i_2, \ldots, i_m \qquad \text{and} \qquad j_1, j_2, \ldots, j_m$$

then $(-1)^{i_1+i_2+\cdots+i_m+j_1+j_2+\cdots+j_m}$ times the complementary minor of $|M|$ is called the **algebraic complement** of $|M|$ The first-order minors of $|A|$ are, of course, just the elements of $|A|$ Their complementary minors are customarily referred to simply as **minors,** and their algebraic complements are almost universally referred to as **cofactors.** We shall denote the minor of the element a_{ij} by the symbol M_{ij} and its cofactor by the symbol A_{ij}; thus,

$$A_{ij} = (-1)^{i+j} M_{ij}$$

Similarly, we shall use the symbols $M_{ij,kl}$ and $A_{ij,kl}$ to denote, respectively, the complementary minor and the algebraic complement of the second-order minor contained in the ith and jth rows and the kth and lth columns of a determinant $|A|$, thus,

$$A_{ij,kl} = (-1)^{i+j+k+l} M_{ij,kl}$$

The generalization of this notation is obvious.

EXAMPLE 1

In the fifth-order determinant

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{vmatrix}$$

the minor of the element a_{43} is the fourth-order determinant formed by the elements which remain when the fourth row and third column are deleted from $|A|$, namely,

$$M_{43} = \begin{vmatrix} a_{11} & a_{12} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{34} & a_{35} \\ a_{51} & a_{52} & a_{54} & a_{55} \end{vmatrix}$$

The cofactor A_{43} of the element a_{43} is equal to this minor times $(-1)^{4+3}$, i e.,

$$A_{43} = -M_{43}$$

Similarly, the complementary minor of the second-order minor

$$\begin{vmatrix} a_{31} & a_{34} \\ a_{51} & a_{54} \end{vmatrix}$$

contained in the third and fifth rows and the first and fourth columns of $|A|$ is the third-order determinant formed by the elements which remain when these rows and columns are deleted from $|A|$

$$M_{35,14} = \begin{vmatrix} a_{12} & a_{13} & a_{15} \\ a_{22} & a_{23} & a_{25} \\ a_{42} & a_{43} & a_{45} \end{vmatrix}$$

The algebraic complement $A_{35,14}$ of the given second-order minor is equal to the complementary minor $M_{35,14}$ times $(-1)^{3+5+1+4}$, i e ,

$$A_{35,14} = -M_{35,14}$$

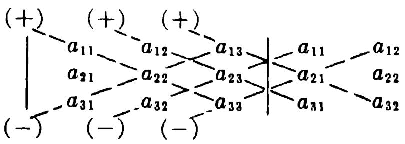
For a second-order determinant we have the definition

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \tag{2}$$

that is, a second-order determinant is equal to the difference between the product of the elements on the principal diagonal and the product of the elements on the other diagonal For a third-order determinant we have the definition

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \tag{3}$$

This expansion can also be obtained by diagonal multiplication, by repeating on the right the first two columns of the determinant and then adding the signed products of the elements on the various diagonals in the resulting array



The diagonal method of writing out a determinant is correct *only* for determinants of the second and third orders, however, and will in general give incorrect results if applied to determinants of higher order.

We are now in a position to give the general definition of a determinant. This can be done in direct fashion, but the result is unsuited to the practical evaluation of determinants; hence, we choose to give an inductive definition:

DEFINITION 1

The determinant

$$|A| = \begin{vmatrix} a_{11} & a_{12} & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot \quad \cdot & a_{2n} \\ \cdot\ \cdot & \cdot & \cdot\ \cdot & \\ a_{n1} & a_{n2} & & a_{nn} \end{vmatrix}$$

is equal to the sum of the products of the elements of any row or column and their respective cofactors; i.e ,

$$|A| = \sum_{j=1}^{n} a_{ij}A_{ij} = \sum_{i=1}^{n} a_{ij}A_{ij}$$

Clearly, this definition makes a determinant of order n depend upon n determinants of order $n - 1$, each of which in turn depends upon $n - 1$ determinants of order $n - 2$, and so on, until finally the expansion involves only second- or third-order determinants which can be written out by the diagonal method. However, before Definition 1 can be accepted and used, it must be sh wn that the same expansion is obtained no matter which row or which column is selected That this is the case is guaranteed by the following theorem

THEOREM 1

If the elements of any row or of any column of a determinant are multiplied by their respective cofactors and then added, the sum is the same for all rows and for all columns

PROOF We shall first prove that the same expansion is obtained no matter which row is chosen To do this we proceed inductively Clearly, the theorem is true when $n = 2$; for, expanding the determinant

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

in terms of the elements of the first row and their cofactors, we get

$$a_{11}(a_{22}) + a_{12}(-a_{21})$$

and, expanding in terms of the elements of the second row and their cofactors we get

$$a_{21}(-a_{12}) + a_{22}(a_{11})$$

and these two expressions are identical. Let us assume, then, that the assertion of the theorem is true for determinants of order $n - 1$, and let us attempt to prove that it is true for determinants of order n. Specifically, let us expand the nth-order determinant

$$|A| = |a_{ij}|$$

in terms of each of two arbitrary rows, say the ith and the jth, and compare the expansions. In doing this it is, of course, no specialization to assume $i < j$.

$$(4a)\qquad \begin{vmatrix} a_{11} & \cdots & \cancel{a_{1k}} & \cdots & a_{1l} & \cdots & a_{1n} \\ \cdots & \cdots & \cancel{\cdots} & \cdots & \cdots & \cdots & \cdots \\ \cancel{a_{i1}} & \cancel{\cdots} & \cancel{a_{ik}} & \cancel{\cdots} & \cancel{a_{il}} & \cancel{\cdots} & \cancel{a_{in}} \\ \cdots & \cdots & \cancel{\cdots} & \cdots & \cdots & \cdots & \cdots \\ a_{j1} & \cdot & \cancel{a_{jk}} & \cdot & \enclose{circle}{a_{jl}} & \cdot & a_{jn} \\ \cdots & \cdots & \cancel{\cdots} & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & \cdots & \cancel{a_{nk}} & \cdots & a_{nl} & \cdots & a_{nn} \end{vmatrix}$$

$$= \begin{vmatrix} a_{11} & \cdots & a_{1k} & \cdots & \cancel{a_{1l}} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cancel{\cdots} & \cdots & \cdots \\ a_{i1} & \cdot & \enclose{circle}{a_{ik}} & \cdots & \cancel{a_{il}} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots & \cancel{\cdots} & \cdots & \cdots \\ \cancel{a_{j1}} & \cancel{\cdots} & \cancel{a_{jk}} & \cancel{\cdots} & \cancel{a_{jl}} & \cancel{\cdots} & \cancel{a_{jn}} \\ \cdots & \cdots & \cdots & \cdots & \cancel{\cdots} & \cdots & \cdots \\ a_{n1} & \cdots & a_{nk} & \cdots & \cancel{a_{nl}} & \cdots & a_{nn} \end{vmatrix} \qquad k < l$$

$$(4b)\qquad \begin{vmatrix} a_{11} & \cdots & a_{1l} & \cdots & \cancel{a_{1k}} & \cdot & a_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cancel{\cdots} & \cdots & \cdots \\ \cancel{a_{i1}} & \cancel{\cdots} & \cancel{a_{il}} & \cancel{\cdots} & \cancel{a_{ik}} & \cancel{\cdots} & \cancel{a_{in}} \\ \cdot & \cdot & & & \cancel{\cdot} & \cdot & \cdot \\ a_{j1} & \cdot & \enclose{circle}{a_{jl}} & & \cancel{a_{jk}} & & a_{jn} \\ \cdots & \cdots & \cdots & \cdots & \cancel{\cdots} & \cdots & \cdots \\ a_{n1} & \cdots & a_{nl} & \cdot & \cancel{a_{nk}} & \cdot & a_{nn} \end{vmatrix}$$

$$= \begin{vmatrix} a_{11} & \cdot & \cancel{a_{1l}} & \cdots & a_{1k} & \cdots & a_{1n} \\ \cdots & \cdots & \cancel{\cdots} & \cdots & \cdots & \cdots & \cdots \\ a_{i1} & \cdot & \cancel{a_{il}} & \cdots & \enclose{circle}{a_{ik}} & \cdots & a_{in} \\ \cdots & \cdots & \cancel{\cdots} & \cdots & \cdots & \cdots & \cdots \\ \cancel{a_{j1}} & \cancel{\cdots} & \cancel{a_{jl}} & \cancel{\cdots} & \cancel{a_{jk}} & \cancel{\cdots} & \cancel{a_{jn}} \\ \cdots & \cdots & \cancel{\cdots} & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & \cdots & \cancel{a_{nl}} & \cdots & a_{nk} & \cdots & a_{nn} \end{vmatrix} \qquad k > l$$

Now, a typical term in the expansion of $|A|$ according to the elements of the ith row is

$$(5)\qquad a_{ik}A_{ik} = (-1)^{i+k}a_{ik}M_{ik}$$

and this contains the only occurrences of a_{ik} in the entire expansion. Moreover, the $(j-1)$st row of M_{ik} contains $n-1$ elements from the jth row of $|A|$, and M_{ik} can legitimately be expanded in terms of these elements, since the hypothesis of our induction is that the theorem in question is true for determinants of order $n-1$. As the typical term in the expansion of M_{ik} according to the elements from the jth row of $|A|$, we therefore have

$$(6)\qquad a_{jl}\cdot(\text{cofactor of } a_{jl} \text{ in } M_{ik}) \qquad l \neq k$$

and this contains the only occurrences of a_{jl} in the expansion of M_{ik}. Hence, substituting the expression (6) into Eq. (5), we find that the expression

$$(7)\qquad (-1)^{i+k}a_{ik}\cdot[a_{jl}\cdot(\text{cofactor of } a_{jl} \text{ in } M_{ik})] \qquad l \neq k$$

contains the only occurrences of the product $a_{ik}a_{jl}$ in the expansion of $|A|$ in terms of the elements of the ith row. In exactly the same way, if we first expand $|A|$ in terms of the elements of the jth row and then expand the minor M_{jl} in terms of the $n - 1$ elements from the ith row of $|A|$ which it contains, we conclude that the only occurrences of the product $a_{ik}a_{jl}$ in the expansion of $|A|$ in terms of the elements of the jth row are contained in the expression

$$(8) \qquad (-1)^{j+l}a_{jl} \cdot [a_{ik} \cdot (\text{cofactor of } a_{ik} \text{ in } M_{jl})] \qquad k \neq l$$

If we can show that (7) and (8) are identical, we will have completed our proof that under the induction hypothesis all row expansions of $|A|$ are the same, since $a_{ik}a_{jl}$ $(k \neq l)$ is the typical product of an element from the ith row and an element from the jth row, and each term in the expansion of $|A|$ according to the ith row or the jth row must contain one and only one such product.

Now, except for the proper power of (-1), both the cofactor of a_{jl} in M_{ik} and the cofactor of a_{ik} in M_{jl} are equal to the determinant of order $n - 2$, say $M_{ij,kl}$, formed by the elements which remain when the ith and jth rows and the kth and lth columns are deleted from $|A|$. There are two possibilities to consider, according as $k < l$ (4a) or $k > l$ (4b), i e , according as the kth column precedes or follows the lth column in the determinant $|A|$ Taking due account of the relative positions of the deleted rows and columns, the proper signs are easily determined by inspection, however, and we have, respectively, in the two cases.

$$\left.\begin{aligned} \text{Cofactor of } a_{jl} \text{ in } M_{ik} &= (-1)^{(j-1)+(l-1)}M_{ij,kl} \\ \text{Cofactor of } a_{ik} \text{ in } M_{jl} &= (-1)^{i+k}M_{ij,kl} \end{aligned}\right\} k < l$$

$$\left.\begin{aligned} \text{Cofactor of } a_{jl} \text{ in } M_{ik} &= (-1)^{(j-1)+l}M_{ij,kl} \\ \text{Cofactor of } a_{ik} \text{ in } M_{jl} &= (-1)^{i+(k-1)}M_{ij,kl} \end{aligned}\right\} k > l$$

Finally, substituting these expressions into (7) and (8), we find that the coefficient of the product $a_{ik}a_{jl}$, as determined by either method of expansion, is

$$(9a) \qquad (-1)^{i+j+k+l}M_{ij,kl} \qquad k < l$$

$$(9b) \qquad (-1)^{i+j+k+l-1}M_{ij,kl} = -(-1)^{i+j+k+l}M_{ij,kl} \qquad k > l$$

In exactly the same way, if we expand $|A|$ in terms of two arbitrary columns, say the kth and the lth, we find that the coefficient of the general product $a_{ik}a_{jl}$ is still given by (9a) and (9b) This proves that, under the induction hypothesis, not only are all column expansions of $|A|$ equal but their common value is the common value of the row expansions of $|A|$ Thus we have completed our proof that *if* the theorem is true for determinants of order $n - 1$, *then* it is true for determinants of order n Since we have already proved it true for row expansions of second-order determinants and could similarly prove it true for column expansions, our induction is complete; Theorem 1 is established; and Definition 1 is unambiguous.

Since the same expression is obtained whether we expand a determinant in terms of the elements of an arbitrary row or an arbitrary column, we have the following obvious consequence of Theorem 1:

THEOREM 2

If $|A|$ is any determinant and if $|B|$ is the determinant whose rows are the columns of $|A|$, then $|A| = |B|$.

The proof of Theorem 1 also provides us with the proof of the following important result.

THEOREM 3

Let any two rows (or columns) be selected from a determinant $|A|$. Then $|A|$ is equal to the sum of the products of all the second-order minors contained in the chosen pair of rows (or columns) each multiplied by its algebraic complement

PROOF Let the chosen rows be the pth and the qth, and, for definiteness, suppose that $p < q$. Now a typical second-order minor from these rows is

$$\begin{vmatrix} a_{pr} & a_{ps} \\ a_{qr} & a_{qs} \end{vmatrix} = a_{pr}a_{qs} - a_{ps}a_{qr} \qquad r < s \tag{10}$$

and to prove the assertion of the theorem it is sufficient to show that the coefficient of this binomial in the expansion of $|A|$ is

$$A_{pq,rs} = (-1)^{p+q+r+s}M_{pq,rs}$$

To do this, we observe first that, from Eq. (9a), with $i = p$, $j = q$, $k = r$, and $l = s$, the coefficient of the product $a_{pr}a_{qs}$ in the expansion of $|A|$ is

$$(-1)^{p+q+r+s}M_{pq,rs}$$

Also, taking $i = p$, $j = q$, $k = s$, and $l = r$ in Eq. (9b), we find that the coefficient of $a_{ps}a_{qr}$ in the expansion of $|A|$ is

$$(-1)^{p+q+r+s}M_{pq,rs}$$

Hence, the expansion of $|A|$ contains the terms

$$a_{pr}a_{qs}[(-1)^{p+q+r+s}M_{pq,rs}] + a_{ps}a_{qr}[-(-1)^{p+q+r+s}M_{pq,rs}]$$

and these are the only occurrences of the products $a_{pr}a_{qs}$ and $a_{ps}a_{qr}$. Finally, from these, by factoring, we obtain

$$(a_{pr}a_{qs} - a_{ps}a_{qr})[(-1)^{p+q+r+s}M_{pq,rs}] = (a_{pr}a_{qs} - a_{ps}a_{qr})A_{pq,rs}$$

which completes the proof of the theorem in the case where two rows are used for the expansion An essentially identical argument establishes the assertion of the theorem when two columns are used.

By a somewhat more involved argument the following generalization of Theorem 3 can be established

THEOREM 4

Let any m rows (or columns) be selected from a determinant $|A|$. Then $|A|$ is equal to the sum of the products of all the mth-order minors contained in the chosen rows (or columns) each multiplied by its algebraic complement.

Both the general result contained in Theorem 4 and the special case $m = 2$ contained in Theorem 3 are usually referred to as **Laplace's expansion.**

EXAMPLE 2

Expand the determinant

$$|A| = \begin{vmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \\ 0 & -1 & 2 & 3 \\ 1 & 6 & 4 & -2 \end{vmatrix}$$

For purposes of illustration we shall obtain the value of this determinant using Definition 1 and also using Theorem 3 According to Definition 1, using the third row because of the presence of the zero element, we have

$$(0)\begin{vmatrix} 2 & 3 & 4 \\ 3 & 2 & 1 \\ 6 & 4 & -2 \end{vmatrix} - (-1)\begin{vmatrix} 1 & 3 & 4 \\ 4 & 2 & 1 \\ 1 & 4 & -2 \end{vmatrix} + (2)\begin{vmatrix} 1 & 2 & 4 \\ 4 & 3 & 1 \\ 1 & 6 & -2 \end{vmatrix} - (3)\begin{vmatrix} 1 & 2 & 3 \\ 4 & 3 & 2 \\ 1 & 6 & 4 \end{vmatrix}$$

or, expanding the third-order determinants by the diagonal method,

$$|A| = 0 + 75 + 180 - 105 = 150$$

Equivalently, applying Theorem 3 in terms of the first two rows, we have

$$\begin{aligned} |A| = {} & \begin{vmatrix} 1 & 2 \\ 4 & 3 \end{vmatrix}\begin{vmatrix} 2 & 3 \\ 4 & -2 \end{vmatrix} - \begin{vmatrix} 1 & 3 \\ 4 & 2 \end{vmatrix}\begin{vmatrix} -1 & 3 \\ 6 & -2 \end{vmatrix} + \begin{vmatrix} 1 & 4 \\ 4 & 1 \end{vmatrix}\begin{vmatrix} -1 & 2 \\ 6 & 4 \end{vmatrix} \\ & + \begin{vmatrix} 2 & 3 \\ 3 & 2 \end{vmatrix}\begin{vmatrix} 0 & 3 \\ 1 & -2 \end{vmatrix} - \begin{vmatrix} 2 & 4 \\ 3 & 1 \end{vmatrix}\begin{vmatrix} 0 & 2 \\ 1 & 4 \end{vmatrix} + \begin{vmatrix} 3 & 4 \\ 2 & 1 \end{vmatrix}\begin{vmatrix} 0 & -1 \\ 1 & 6 \end{vmatrix} \\ = {} & (-5)(-16) - (-10)(-16) + (-15)(-16) + (-5)(-3) - (-10)(-2) + (-5)(1) \\ = {} & 150 \end{aligned}$$

as before

Using Theorems 1 and 3, a number of other theorems can easily be proved In particular, we have the following useful results·

THEOREM 5

If all the elements in any row or in any column of a determinant are zero, the value of the determinant is zero

PROOF If we expand the given determinant, according to Definition 1, in terms of the row or column of zero elements, each term in the expansion contains a zero factor Hence, the entire expansion is zero, as asserted

THEOREM 6

If each element in one row or in one column of a determinant is multiplied by c, the determinant is multiplied by c

PROOF If we expand the given determinant in terms of the row or column whose elements have been multiplied by c, each term in the expansion contains c as a factor If c is then factored from the expansion, the result is just c times the expansion of the original determinant, as asserted.

THEOREM 7

If $|A|$ is any determinant and if $|B|$ is the determinant obtained from $|A|$ by interchanging any two rows or any two columns of $|A|$, then $|B| = -|A|$.

PROOF Let $|A|$ be any determinant, and let $|B|$ be the determinant obtained from $|A|$ by interchanging any two rows (or any two columns) of $|A|$. Now, clearly, if the rows (or the columns) of any second-order determinant are interchanged, the resulting determinant is the negative of the original one. Hence, if $|B|$ is expanded in terms of the two rows (or columns) which were interchanged, it follows that each second-order minor occurring as a factor of a term in this expansion is the negative of the corresponding second-order minor from the corresponding pair of rows (or columns) in $|A|$ Therefore, each term in the expansion of $|B|$ is the negative of the corresponding term in the expansion of $|A|$ based on the same two rows (or columns). Thus $|B| = -|A|$, as asserted

THEOREM 8

If corresponding elements of two rows or of two columns of a determinant are proportional, the value of the determinant is zero

PROOF Clearly, any determinant of the second order whose rows or columns are proportional is zero. Hence if we expand the given determinant, according to Theorem 3, in terms of the two rows or two columns which are proportional, it follows that each term contains as one factor a second-order minor equal to zero Therefore, the entire expansion is zero, as asserted

THEOREM 9

If the elements in one column of a determinant are expressed as binomials, the determinant can be written as the sum of two determinants, according to the formula

$$\begin{vmatrix} a_{11} & \cdot & a_{1j}+\alpha_{1j} & \cdot & a_{1n} \\ a_{21} & \cdot & a_{2j}+\alpha_{2j} & \cdot\ \cdot & a_{2n} \\ \cdots & \cdot & \cdots & \cdot & \cdot \\ a_{n1} & \cdot & a_{nj}+\alpha_{nj} & & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & \cdot & a_{1j} & \cdot\ \cdot & a_{1n} \\ a_{21} & & a_{2j} & & a_{2n} \\ & \cdot & & \cdot & \\ a_{n1} & & a_{nj} & & a_{nn} \end{vmatrix}$$

$$+ \begin{vmatrix} a_{11} & & \alpha_{1j} & \cdot & a_{1n} \\ a_{21} & \cdot & \alpha_{2j} & \cdot\ \cdot\ \cdot & a_{2n} \\ \cdot & \cdot\ \cdot & & \cdot & \cdot \\ a_{n1} & \cdot & \alpha_{nj} & & a_{nn} \end{vmatrix}$$

A similar result holds for a determinant containing a row of elements which are binomials.

PROOF If we expand the given determinant, according to Definition 1, in terms of the column which contains the binomial elements, we obtain

$$\sum_{i=1}^{n} (a_{ij} + \alpha_{ij})A_{ij} = \sum_{i=1}^{n} a_{ij}A_{ij} + \sum_{i=1}^{n} \alpha_{ij}A_{ij}$$

Since the sums on the right are, respectively, the expansions for the determinants appearing on the right side of the formula in the theorem, the theorem is established.

THEOREM 10

The value of a determinant is unchanged if the elements of any row (or column) are modified by adding to them the same multiple of the corresponding elements in any other row (or column)

PROOF If we apply Theorem 9 to the determinant resulting from the given row (or column) modification, we obtain two determinants, one of which is the original determinant and the other of which contains two proportional rows (or columns) By Theorem 8, the second determinant is equal to zero, and the theorem is established.

Theorem 10 is very useful in the practical expansion of determinants, for, by its repeated application, one can reduce to zero a number of the elements in a chosen row (or column) of the given determinant Then, when the determinant is expanded in terms of this row (or column) most of the products involved will be zero and the computation will be appreciably shortened.

EXAMPLE 3

Find the value of the determinant

$$\begin{vmatrix} 3 & 1 & -1 & 2 & 1 \\ 0 & 3 & 1 & 4 & 2 \\ 1 & 4 & 2 & 3 & 1 \\ 5 & -1 & -3 & 2 & 5 \\ -1 & 1 & 2 & 3 & 2 \end{vmatrix}$$

Here, in an attempt to introduce as many zeros as possible into some row, let us add the third column to the second and to the fifth, and let us add twice the third column to the fourth and 3 times the third column to the first This gives the new but equal determinant

$$\begin{vmatrix} 0 & 0 & -1 & 0 & 0 \\ 3 & 4 & 1 & 6 & 3 \\ 7 & 6 & 2 & 7 & 3 \\ -4 & -4 & -3 & -4 & 2 \\ 5 & 3 & 2 & 7 & 4 \end{vmatrix}$$

Expanding this in terms of the first row, according to Definition 1, we have

$$(-1)(-1)^{1+3}\begin{vmatrix} 3 & 4 & 6 & 3 \\ 7 & 6 & 7 & 3 \\ -4 & -4 & -4 & 2 \\ 5 & 3 & 7 & 4 \end{vmatrix}$$

Now, adding twice the last column to each of the first three, we obtain the equal determinant

$$-\begin{vmatrix} 9 & 10 & 12 & 3 \\ 13 & 12 & 13 & 3 \\ 0 & 0 & 0 & 2 \\ 13 & 11 & 15 & 4 \end{vmatrix}$$

or, expanding in terms of the third row,

$$-(2)(-1)^{3+4}\begin{vmatrix} 9 & 10 & 12 \\ 13 & 12 & 13 \\ 13 & 11 & 15 \end{vmatrix}$$

We can now simplify this by further row or column manipulations, or, since it is of the third order, we can expand it by the diagonal method The result is -166.

THEOREM 11

The sum of the products formed by multiplying the elements of one row (or column) of a determinant by the cofactors of the corresponding elements of another row (or column) is zero.

PROOF Let $|A| = |a_{ij}|$ be the given determinant, and let the elements of some row of $|A|$, say the ith, be multiplied by the cofactors of the corresponding elements in some other row, say the jth, giving the sum

$$\sum_{k=1}^{n} a_{ik}A_{jk}$$

Clearly, according to Definition 1, this can be thought of as the expansion of a determinant whose jth row consists of the elements

$$a_{i1} \qquad a_{i2} \qquad \cdots \qquad a_{in}$$

and whose other rows are identical with the corresponding rows in $|A|$. In this new determinant the ith and jth rows are therefore the same, and, hence, by Theorem 8, the determinant is equal to zero. A similar argument leads to the same conclusion if the elements in some column of $|A|$ are multiplied by the cofactors of the corresponding elements in some other column of $|A|$. Thus the theorem is established.

Combining Definition 1 and Theorem 11, we have the following useful result.

COROLLARY 1

If A_{ij} is the cofactor of the element a_{ij} in the determinant $|A| = |a_{ij}|$, then

$$\sum_{k=1}^{n} a_{ik}A_{jk} = \begin{cases} 0 & i \neq j \\ |A| & i = j \end{cases} \qquad \text{and} \qquad \sum_{i=1}^{n} a_{ik}A_{il} = \begin{cases} 0 & k \neq l \\ |A| & k = l \end{cases}$$

EXAMPLE 4

If we take the elements of the first row of the determinant

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

and multiply them by the cofactors of the corresponding elements in the third row, say, we obtain the sum

$$a_{11}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} - a_{12}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} + a_{13}\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

and this is clearly the expansion of the determinant

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \end{vmatrix}$$

according to the third row. Since this determinant has two identical rows, it vanishes identically.

Results such as those of Corollary 1 are often stated more compactly in terms of what is known as the **Kronecker delta,*** usually written δ_{ij}, or sometimes δ_j^i, and defined to be 0 or 1 according as $i \neq j$ or $i = j$. Using the Kronecker delta, the assertions of Corollary 1 can be written in the simpler form

---

* Named for the German mathematician Leopold Kronecker (1823–1891).

$$\sum_{k=1}^{n} a_{ik}A_{jk} = |A|\delta_{ij} \tag{11}$$

$$\sum_{i=1}^{n} a_{ik}A_{il} = |A|\delta_{kl} \tag{12}$$

THEOREM 12

The product of two determinants of the same order is a determinant of the same order in which the element in the ith row and jth column is the sum of the products of corresponding elements in the ith row of the first determinant and the jth column of the second determinant

PROOF For simplicity we shall prove this theorem only for determinants of the second order, although for these direct verification is easier and more natural than the method we shall actually use. The virtue of our proof is that it can be extended immediately to the general case of determinants of any order We begin by observing that if

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \quad \text{and} \quad |B| = \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix}$$

then, by Theorem 3,

$$|A|\ |B| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ c_{11} & c_{12} & b_{11} & b_{12} \\ c_{21} & c_{22} & b_{21} & b_{22} \end{vmatrix}$$

where c_{11}, c_{12}, c_{21}, and c_{22} are completely arbitrary. In particular, it is convenient to take $c_{11} = c_{22} = -1$ and $c_{12} = c_{21} = 0$, so that we have

$$|A|\ |B| = \begin{vmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ -1 & 0 & b_{11} & b_{12} \\ 0 & -1 & b_{21} & b_{22} \end{vmatrix}$$

Now if we multiply the elements of the first column by b_{11} and the elements of the second column by b_{21} and then add them to the corresponding elements of the third column, we obtain, by Theorem 10, the equal determinant

$$|A|\ |B| = \begin{vmatrix} a_{11} & a_{12} & a_{11}b_{11} + a_{12}b_{21} & 0 \\ a_{21} & a_{22} & a_{21}b_{11} + a_{22}b_{21} & 0 \\ -1 & 0 & 0 & b_{12} \\ 0 & -1 & 0 & b_{22} \end{vmatrix}$$

In the same way, if we multiply the elements of the first column by b_{12} and the elements of the second column by b_{22} and then add them to the corresponding elements of the fourth column, we obtain from the last determinant the equal determinant

$$|A|\ |B| = \begin{vmatrix} a_{11} & a_{12} & a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21} & a_{22} & a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{vmatrix}$$

If we now expand this determinant by Theorem 3, applied to the last two rows, we obtain

$$|A| \cdot |B| = \begin{vmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{vmatrix}$$

which is the result asserted by the theorem.

EXERCISES

1 Find the value of each of the following determinants

a $\begin{vmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 2 & 1 \\ 4 & 3 & 1 & 2 \end{vmatrix}$ **b** $\begin{vmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 1 & 2 \end{vmatrix}$ **c** $\begin{vmatrix} 0 & 1 & 2 & 3 \\ -1 & 0 & 1 & 2 \\ -2 & -1 & 0 & 3 \\ -3 & -2 & -3 & 0 \end{vmatrix}$

2 **a** Find the value(s) of a, if any, for which the diagonal method of expansion yields the correct value for the determinant

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ -1 & 2 & 0 & 3 \\ 2 & 0 & c & 1 \\ 1 & 4 & -9 & a \end{vmatrix}$$

b Show that there is no value of a for which the diagonal method of expansion yields the correct value for the determinant

$$\begin{vmatrix} 1 & 2 & 1 & 1 \\ -1 & 1 & 3 & 2 \\ -1 & 3 & 9 & 1 \\ 2 & 1 & 1 & a \end{vmatrix}$$

3 Show that the number of terms in the expansion of a general determinant of order n is $n!$

4 If $|A| = |a_{ij}|$ is a determinant of order n with the property that $a_{ij} = a_{ji}$ for all values of i and j such that $1 \leq i,j \leq n$, prove that $|A| = (-1)^n |A|$ What further conclusion can be drawn if n is odd? (Hint Use Theorems 2 and 6)

5 Prove that

$$\begin{vmatrix} 1 + a_1 & a_2 & a_3 & & a_n \\ a_1 & 1 + a_2 & a_3 & & a_n \\ a_1 & a_2 & 1 + a_3 & \cdot \quad \cdot & a_n \\ \cdot & & & & \\ a_1 & a_2 & a_3 & & 1 + a_n \end{vmatrix} = 1 + a_1 + a_2 + a_3 + \qquad + a_n$$

6 If D_n is the nth-order determinant

$$\begin{vmatrix} 1 + x^2 & x & 0 & & 0 & 0 \\ x & 1 + x^2 & x & & 0 & 0 \\ 0 & x & 1 + x^2 & \cdot & 0 & 0 \\ & \cdot & \cdot\ \cdot & \cdot & \cdot & \\ 0 & 0 & 0 & \cdot\ \cdot & 1 + x^2 & x \\ 0 & 0 & 0 & \cdot & x & 1 + x^2 \end{vmatrix}$$

show that $D_n = (1 + x^2)D_{n-1} - x^2D_{n-2}$ Using this relation, determine the value of D_{10} if $x = 1$; if $x = -1$ Is the value of D_n independent of x?

7 If D_n is the nth-order determinant in which each element on the principal diagonal is a, each element immediately above the principal diagonal is b, each element immediately below the principal diagonal is c, and all other elements are zero, obtain a recurrence relation expressing D_n in terms of D_{n-1} and D_{n-2}. Use this relation to infer the value of D_n if $a = 3$, $b = 2$, $c = 1$

8 Show that the nth-order determinant

$$\begin{vmatrix} a & b & \cdot & & b & b \\ b & a & & \cdot\ \cdot & b & b \\ \cdot & & & \cdot\ \cdot & & \\ b & b & & \cdot & a & b \\ b & b & & \cdot & b & a \end{vmatrix}$$

is equal to $(a - b)^{n-1}[a + (n - 1)b]$

9 Show that

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ a_1 & a_2 & a_3 & a_4 \\ a_1^2 & a_2^2 & a_3^2 & a_4^2 \\ a_1^3 & a_2^3 & a_3^3 & a_4^3 \end{vmatrix} = (a_1 - a_2)(a_1 - a_3)(a_1 - a_4)(a_2 - a_3)(a_2 - a_4)(a_3 - a_4)$$

What is the generalization of this result to determinants of order n?

10 If $p_1, p_2, \ldots, p_n$ are polynomials and if $x_1, x_2, \ldots, x_n$ are variables, show that the nth-order determinant

$$\begin{vmatrix} p_1(x_1) & p_1(x_2) & \cdot & p_1(x_n) \\ p_2(x_1) & p_2(x_2) & & p_2(x_n) \\ & & & \\ p_n(x_1) & p_n(x_2) & & p_n(x_n) \end{vmatrix}$$

is divisible by $\displaystyle\prod_{1 \le i < j \le n} (x_i - x_j)$.

11 Prove the following generalization of Theorem 12. The product of two determinants of the same order is another determinant of the same order in which the element in the ith row and jth column is the sum of the products of corresponding elements in the ith row *or* column of the first determinant and the jth row *or* column of the second determinant, a consistent choice of row or column being maintained for all values of i and j. (Hint. Use Theorem 2.)

12 **a** Show that $|A| = \begin{vmatrix} b^2 + ac & bc & c^2 \\ ab & ac & bc \\ a^2 & ab & b^2 + ac \end{vmatrix} = 4a^2b^2c^2$

$\left(\text{Hint. Verify first that } |A| = \begin{vmatrix} b & c & 0 \\ a & 0 & c \\ 0 & a & b \end{vmatrix}^2\right)$

b Find the value of the determinant $\begin{vmatrix} b^2 + c^2 & ab & ca \\ ab & c^2 + a^2 & bc \\ ca & bc & a^2 + b^2 \end{vmatrix}$

13 If $|A|$ is the nth-order determinant

$$\begin{vmatrix} a_{11} & a_{12} & & a_{1n} \\ a_{21} & a_{22} & \cdot & a_{2n} \\ & & & \\ a_{n1} & a_{n2} & & a_{nn} \end{vmatrix} \quad \text{show that} \quad \begin{vmatrix} a_{11} & a_{12} & & a_{1n} & x_1 \\ a_{21} & a_{22} & \cdot & a_{2n} & x_2 \\ \cdot & & \cdot\ \cdot\ \cdot & & \\ a_{n1} & a_{n2} & \cdot & a_{nn} & x_n \\ y_1 & y_2 & \cdot\ \cdot & y_n & 0 \end{vmatrix} = -\sum_{i,j=1}^{n} A_{ij}x_iy_j$$

where A_{ij} is the cofactor of a_{ij} in $|A|$

14 Show that the area of the triangle whose vertices are the points (x_1,y_1), (x_2,y_2), (x_3,y_3) is given by the formula

$$A = \pm\frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$$

where the plus or the minus sign is chosen according as the vertices of the triangle are numbered consecutively in the counterclockwise or the clockwise direction.

15 If P_1 (x_1,y_1), P_2 (x_2,y_2), and P_3 (x_3,y_3) are three points, no two of which lie on the same vertical line, show that the equation of the parabola of the family $y = a + bx + cx^2$ which passes through P_1, P_2, and P_3 can be written in the form

$$\begin{vmatrix} y & 1 & x & x^2 \\ y_1 & 1 & x_1 & x_1^2 \\ y_2 & 1 & x_2 & x_2^2 \\ y_3 & 1 & x_3 & x_3^2 \end{vmatrix} = 0$$

Is this result correct if P_1, P_2, P_3 are collinear?

16 Show that the equation of the circle which passes through the three points P_1 (x_1,y_1), P_2 (x_2,y_2), and P_3 (x_3,y_3) can be written in the form

$$\begin{vmatrix} x^2 & y^2 & x & y & 1 \\ x_1^2 & y_1^2 & x_1 & y_1 & 1 \\ x_2^2 & y_2^2 & x_2 & y_2 & 1 \\ x_3^2 & y_3^2 & x_3 & y_3 & 1 \end{vmatrix} = 0$$

Is this result correct if the three points are collinear?

17 **a** If

$$\begin{aligned} l_1 \quad & a_{11}x + a_{12}y + a_{13} = 0 \\ l_2 \quad & a_{21}x + a_{22}y + a_{23} = 0 \\ l_3 \quad & a_{31}x + a_{32}y + a_{33} = 0 \end{aligned}$$

are three lines no two of which are parallel, show that l_1, l_2, and l_3 are concurrent if and only if

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = 0$$

b Show that the area of the triangle determined by the lines l_1, l_2, and l_3 is equal to the absolute value of the expression

$$\frac{1}{A_{11}A_{22}A_{33}}\begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix}$$

where A_{ij} is the cofactor of a_{ij} in $|A|$

18 If $|A| = |a_{ij}|$, show that $\dfrac{\partial|A|}{\partial a_{ij}} = A_{ij}$

19 If each element of a determinant $|A|$ of order n is a differentiable function of t, show that the derivative of $|A|$ with respect to t is equal to the sum of n determinants, the ith one of which is identical with $|A|$ except for the ith row which consists of the derivatives of the elements of the ith row of $|A|$ (Hint Proceed inductively, as in the proof of Theorem 1)

20 If $f_1, f_2, \ldots, f_n$ are suitably differentiable functions of t, show that

$$\frac{d}{dt}\begin{vmatrix} f_1 & \cdots & f_n \\ f_1' & \cdot & f_n' \\ \cdots & & \cdot \\ f_1^{(n-2)} & \cdot\ \cdot & f_n^{(n-2)} \\ f_1^{(n-1)} & \cdots & f_n^{(n-1)} \end{vmatrix} = \begin{vmatrix} f_1 & \cdot & f_n \\ f_1' & \cdot & f_n' \\ & \cdot & \cdot \\ f_1^{(n-2)} & \cdot & f_n^{(n-2)} \\ f_1^{(n)} & \cdots & f_n^{(n)} \end{vmatrix}$$

(**Hint: Use the result of Exercise 19.**)

10.2

Elementary properties of matrices

Closely associated with determinants, yet significantly different and much more fundamental, are the mathematical objects known as **matrices**:

DEFINITION 1

An $m \times n$ or (m,n) matrix is a rectangular array of quantities arranged in m rows and n columns.

When there is no possibility of confusion, matrices are often represented by single capital letters More commonly, however, they are represented by displaying some or all of the constituent quantities between double vertical bars,* thus,

$$A = \|a_{ij}\| = \left\| \begin{matrix} a_{11} & a_{12} & \cdot & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot & \cdot \\ a_{m1} & a_{m2} & & a_{mn} \end{matrix} \right\|$$

Two matrices $A = \|a_{ij}\|$ and $B = \|b_{ij}\|$ are **equal** if and only if they are identical, that is, if and only if they contain the same number of rows and the same number of columns, and $a_{ij} = b_{ij}$ r all values of i and j

A matrix consisting of a single column is called a **column matrix.** A matrix consisting of a single row is called a **row matrix.** Both column matrices and row matrices are often referred to as **vectors.** The (n,m) matrix obtained from a given (m,n) matrix A by interchanging its rows and columns is called the **transpose** of A We shall denote the transpose of a matrix A by the symbol A^T, although some writers use the symbol A' or the symbol $\tilde{A}$. A matrix with the same number of rows and columns is called a **square matrix.** The determinant whose array of elements is identical with the array of elements of a square matrix is called the **determinant of the matrix.** A square matrix in which every element below the principal diagonal is zero is said to be **upper triangular.** A square matrix in which every element above the principal diagonal is zero is said to be **lower triangular.** A square matrix in which each element not on the principal diagonal is zero is called a **diagonal matrix.** Diagonal matrices are sometimes denoted by the symbol

$$\left\| \begin{matrix} a_{11} & & \\ & \ddots & \text{\Large O} \\ \text{\Large O} & \ddots & \\ & & a_{nn} \end{matrix} \right\|$$

---

* Some writers use square brackets or parentheses instead of double bars.

A diagonal matrix in which each diagonal element is 1 is called a **unit matrix.** A unit matrix is usually denoted by the symbol I, or, more specifically, by the symbol I_n if it is a unit matrix of order n. A matrix in which every element is zero is called a **null matrix** or a **zero matrix,** and is denoted by the symbol O

If A is an (m,n) matrix and if k and l are integers such that $0 < k \leqq m$ and $0 < l \leqq n$, then the array of elements common to any k rows and any l columns of A is called a (k,l) **submatrix** of A. If A is a square matrix, any square submatrix of A whose principal diagonal is a part of the principal diagonal of A is called a **principal submatrix** of A A principal submatrix of a square matrix A is thus a submatrix of rows and columns with the same indices. The determinants of the square submatrices of any matrix A are called the **minors** of A. The determinant of any principal submatrix of a square matrix A is called a **principal minor** of A

Most matrices which occur in elementary applications have the property that all their elements are real. However, there are important applications, especially in mathematical physics and quantum mechanics, which involve matrices whose elements are not real. For this reason we shall introduce certain definitions and later on state certain fundamental theorems in a form appropriate to matrices whose elements may be general complex quantities

Recalling that the conjugate of a complex number $z = x + iy$ is the complex number $\bar{z} = x - iy$, we say that the **conjugate** of a matrix A is the matrix $\bar{A}$ whose elements are, respectively, the conjugates of the elements of A. Clearly, a matrix is **real,** i.e., contains only real elements, if and only if A and $\bar{A}$ are the same matrix. Similarly, a matrix A is **imaginary,** i e , contains only elements which are pure imaginaries, if and only if $A = -\bar{A}$ The transpose of the conjugate of a matrix A is called the **associate** of A.

A matrix equal to its transpose, i e , a square matrix such that $a_{ij} = a_{ji}$ for $1 \leqq i,j \leqq n$, is said to be **symmetric.** A matrix equal to the negative of its transpose, i.e , a square matrix such that $a_{ij} = -a_{ji}$ and in which, therefore, $a_{ii} = 0$, is said to be **skew-symmetric.** A matrix equal to its associate, i.e , a matrix A such that $A = \bar{A}^T$, is said to be **hermitian.*** A matrix A such that $A = -\bar{A}^T$ is said to be **skew-hermitian.** Clearly, *a real symmetric matrix is just a hermitian matrix which is real,* and *a real skew-symmetric matrix is just a skew-hermitian matrix which is real.* Thus, in particular, any result true for hermitian matrices is automatically true for real symmetric matrices For this reason, although real symmetric matrices are of fundamental importance in most of the applications of matrices in this book, we shall as far as possible state our theorems in terms of hermitian matrices.

---

* Named for the French mathematician Charles Hermite (1822–1905).

The concept of a matrix is essentially simpler than the concept of a determinant. For, whereas a matrix is just a collection of elements arranged in a particular way, a determinant is a rather complicated function of the elements in a given set. In particular, with every square matrix there is associated a determinant, namely, the determinant whose elements are respectively equal to the elements of the matrix. Thus determinants of order n and $n \times n$ matrices bear to each other the familiar relation of dependent and independent variable, respectively, and it is appropriate to speak of a determinant as a function of a square matrix.

Examples of matrices can be found in many fields. For instance, if each of m students were given a battery of n different tests, the resulting scores would very probably be displayed in a table containing m rows—one for each student—and n columns—one for each test. The resulting array would, of course, be an (m,n) matrix in which the general element a_{ij} was the score which the ith student made on the jth test. Matrices of this sort are of fundamental importance in the branch of mathematical psychology known as *factor analysis*. Similarly, if we had an electrical network containing n branches, we might, either experimentally or analytically, determine the current which would flow in the ith branch as a result of inserting a unit voltage in the jth branch. A tabular array of these quantities would also constitute a matrix—the so-called **admittance matrix** which is of fundamental importance in the theory of electrical circuits. Still another example of a matrix is provided by an array of **transition probabilities**: Suppose that a system S can exist in any one of n states, say S_1, S_2, $\dots$, S_n, and that the probability of the system passing from the state S_i to the state S_j by some well-defined random process is p_{ij}. Clearly, the p_{ij}'s can be displayed as an (n,n) matrix. Such matrices are of great importance in the theory of probability and its physical applications.

Since, as we pointed out above, both row and column matrices are called vectors, we have, in effect, accepted the following definition:

DEFINITION 2

An n-dimensional vector is an ordered set of n quantities.

This use of the word *vector* requires explanation, since, at first glance, it appears to be somewhat at variance with the familiar usage of elementary physics. In physics, a vector quantity is a quantity, such as a velocity or a force, which possesses both magnitude and direction and, hence, can be represented by a directed line segment. However, it is clear that such a quantity is uniquely determined by its components in the directions of the three coordinate axes, and conversely. Hence, it can be uniquely associated with an ordered set of three quantities, i.e., with either a row matrix or a column matrix containing three elements. A vector quantity in the physical sense is thus an example of a vector in the matric sense. However, the matric sense includes vectors other than physical vectors. In particular, any set of values of $x_1, x_2, \ldots, x_n$ which satisfies a system of n linear equations in n variables can be thought of as a vector, and in our work we shall often speak of the **solution vectors** of such systems

By defining appropriately the addition and multiplication of matrices, an algebra of matrices can be developed As we shall soon see, this is quite different from ordinary algebra, and for this reason, it is convenient to have a term to denote collectively those quantities, such as the variables and constants of our work up to this point, which obey the familiar laws of elementary algebra. These we shall henceforth refer to as **scalars.** For matrices, in contrast to scalars, then, we have the following definitions and rules of operation.

The **sum** or **difference** of two matrices A and B having the same number of rows and the same number of columns is the matrix $A \pm B$ whose elements are the sums or differences of the respective elements of A and B. Obviously, if addition is commutative for the elements of A and B, it is also commutative for A and B themselves, and we have $A + B = B + A$ Similarly, if addition is associative for the elements of the matrices A, B, and C, it is associative for A, B, and C, and we have $A + (B + C) = (A + B) + C$ Addition and subtraction are not defined for matrices which do not have the same number of rows and the same number of columns

The **product of a matrix** A **and a scalar** k is the matrix $kA = Ak$ whose elements are the elements of A each multiplied by k

The **scalar product of two vectors** having the same number of elements, or **components,** is the sum of the products of corresponding components of the two vectors The scalar product of two vectors X and Y is also referred to as the **inner product** or **dot product** of X and Y and is often denoted by the symbol $X \cdot Y$. Obviously, $X \cdot Y = Y \cdot X$.

The coordinates of a point in three dimensions, say $P\colon (x_1,x_2,x_3)$, form a vector $X = \|x_1\ x_2\ x_3\|$ in the matric sense, which is completely equivalent to the directed line segment from the origin O to the point P, thought of as a vector in the physical or geometric sense. Now from analytic geometry we know that the square of the length of the segment $\overline{OP}$ is given by the formula

$$(OP)^2 = x_1^2 + x_2^2 + x_3^2$$

But this is simply the scalar product $X \cdot X$ of the vector X with itself. Hence, by analogy, the **length** or **absolute value** of any vector

$$X = \|x_1\ x_2\ \cdot\ x_n\|$$

is defined to be the square root of the scalar product

$$X\ X = \sum_{i=1}^{n} x_i^2$$

A vector X with the property that

$$X \cdot X = \sum_{i=1}^{n} x_i^2 = 1$$

is called a **unit vector.**

From analytic geometry we also know that with every line l, there is associated a set of ordered triples

$$(kl_1,kl_2,kl_3) \qquad k \neq 0;\ l_1,l_2,l_3 \text{ not all zero}$$

known as the *direction numbers* of the line Moreover, if (l_1,l_2,l_3) and (m_1,m_2,m_3) are, respectively, direction numbers of the lines l and m, then l and m are parallel if and only if the sets (l_1,l_2,l_3) and (m_1,m_2,m_3) are proportional Since the sets (l_1,l_2,l_3) and (m_1,m_2,m_3) can obviously be thought of as vectors

$$L = \|l_1\ l_2\ l_3\| \qquad \text{and} \qquad M = \|m_1\ m_2\ m_3\|$$

it is natural to extend these ideas to vectors in general by saying, as a matter of definition, that **two nonzero vectors**

$$X = \|x_1\ x_2\ \cdot\ x_n\| \qquad \text{and} \qquad Y = \|y_1\ y_2\ \cdot\ y_n\|$$

have the same direction if and only if their components are proportional.

It is also a well-known fact of analytic geometry that, if (l_1,l_2,l_3) and (m_1,m_2,m_3) are, respectively, direction numbers of the lines l and m, then l and m are perpendicular if and only if

$$l_1m_1 + l_2m_2 + l_3m_3 = 0$$

But this is simply the condition that the scalar product $L{\cdot}M$ of the two vectors L and M be equal to zero. Hence, by analogy with this result, we agree that **two nonzero vectors**

$$X = \|x_1\ x_2\ \ x_n\| \qquad \text{and} \qquad Y = \|y_1\ y_2\ \cdot\cdot\ y_n\|$$

will be called perpendicular, or orthogonal, if and only if they satisfy the condition

$$X \cdot Y = \sum_{i=1}^{n} x_iy_i = 0$$

This extended concept of orthogonality is of fundamental importance in many applications of matrices

Two matrices A and B are said to be **conformable in the order** AB if and only if the number of columns in A is equal to the number of rows in B. In other words, if A is an (m,n) matrix and if B is a (p,q) matrix, A and B are conformable in the order AB if and only if $n = p$. With the idea of conformable matrices introduced, we are now in a position to define the important notion of the **product** of two matrices:

DEFINITION 3

If A is a (p,q) matrix and if B is a (q,r) matrix, so that A and B are conformable in the order AB, the product $C = AB$ is the (p,r) matrix in which the element c_{ij} in the ith row and jth column is the scalar product of the ith row vector of A and the jth column vector of B, i.e., $C = AB$ is the matrix for which $c_{ij} = \sum_{k=1}^{q} a_{ik}b_{kj}$

Multiplication is not defined for matrices that are not conformable

EXAMPLE 1

$$\begin{Vmatrix} 2 & 3 \\ 1 & -1 \\ 0 & 4 \end{Vmatrix} \cdot \begin{Vmatrix} 5 & -2 & 4 & 7 \\ -6 & 1 & -3 & 0 \end{Vmatrix}$$

$$= \begin{Vmatrix} (2)(5)+(3)(-6) & (2)(-2)+(3)(1) & (2)(4)+(3)(-3) & (2)(7)+(3)(0) \\ (1)(5)+(-1)(-6) & (1)(-2)+(-1)(1) & (1)(4)+(-1)(-3) & (1)(7)+(-1)(0) \\ (0)(5)+(4)(-6) & (0)(-2)+(4)(1) & (0)(4)+(4)(-3) & (0)(7)+(4)(0) \end{Vmatrix}$$

$$= \begin{Vmatrix} -8 & -1 & -1 & 14 \\ 11 & -3 & 7 & 7 \\ -24 & 4 & -12 & 0 \end{Vmatrix}$$

From Theorem 12, Sec. 10 1, and Definition 3, it is clear that the way in which we have *defined* the product of two matrices is precisely the way in which we *proved* that the product of two determinants can be formed. Hence, we have the following important result:

THEOREM 1

If A and B are square matrices of the same order, the determinant of the product AB is equal to the product of the determinant of A and the determinant of B

For the multiplication of matrices we have the following important theorems:

THEOREM 2

For suitably conformable matrices, multiplication is associative, i.e.,

$$A(BC) = (AB)C$$

PROOF Let A be an (m,n) matrix, B an (n,p) matrix, and C a (p,q) matrix For convenience, let $BC = D$, $AB = E$, and let

$$A(BC) = AD = F \qquad \text{and} \qquad (AB)C = EC = G$$

Now the element f_{ij} in the ith row and jth column of the matrix $F = AD$ is, by Definition 3,

$$f_{ij} = \sum_{k=1}^{n} a_{ik}d_{kj}$$

Moreover, since $D = BC$, we also have, by Definition 3,

$$d_{kj} = \sum_{l=1}^{p} b_{kl}c_{lj}$$

Hence, substituting,

$$f_{ij} = \sum_{k=1}^{n} a_{ik}\Big(\sum_{l=1}^{p} b_{kl}c_{lj}\Big) = \sum_{k=1}^{n}\sum_{l=1}^{p} a_{ik}b_{kl}c_{lj} \tag{1}$$

where the last step follows from the fact that a_{ik} is independent of the index l of the inner summation and, hence, can be moved across the inner summation sign

Similarly, the element g_{ij} in the ith row and jth column of the matrix $G = EC$ is

$$g_{ij} = \sum_{l=1}^{p} e_{il}c_{lj} \qquad \text{where} \quad e_{il} = \sum_{k=1}^{n} a_{ik}b_{kl}$$

Hence, substituting,

$$g_{ij} = \sum_{l=1}^{p}\Big(\sum_{k=1}^{n} a_{ik}b_{kl}\Big)c_{lj} = \sum_{l=1}^{p}\sum_{k=1}^{n} a_{ik}b_{kl}c_{lj}$$

Finally, interchanging the order of summation in the last double sum, we have

$$g_{ij} = \sum_{k=1}^{n}\sum_{l=1}^{p} a_{ik}b_{kl}c_{lj} \tag{2}$$

From (1) and (2) it is clear that $f_{ij} = g_{ij}$ for all values of i and j Hence, $F = G$, i e.,

$$A(BC) = (AB)C \qquad \text{as asserted}$$

It is interesting and helpful to note that the type symbol of the product of a series of conformable matrices can be obtained by "contracting" the type symbols of the factors by canceling the common interior indices:

$$(m_1,\cancel{m_2})(\cancel{m_2},\cancel{m_3}) \qquad (\cancel{m_{k-2}},\cancel{m_{k-1}})(\cancel{m_{k-1}},m_k) \longrightarrow (m_1,m_k)$$

EXAMPLE 2

$$\begin{Vmatrix} 3 & 4 \\ 2 & 1 \end{Vmatrix}\left(\begin{Vmatrix} 1 & 2 \\ 2 & 5 \end{Vmatrix}\begin{Vmatrix} 2 & -1 \\ 0 & 3 \end{Vmatrix}\right) = \begin{Vmatrix} 3 & 4 \\ 2 & 1 \end{Vmatrix}\begin{Vmatrix} 2 & 5 \\ 4 & 13 \end{Vmatrix} = \begin{Vmatrix} 22 & 67 \\ 8 & 23 \end{Vmatrix}$$

$$\left(\begin{Vmatrix} 3 & 4 \\ 2 & 1 \end{Vmatrix}\cdot\begin{Vmatrix} 1 & 2 \\ 2 & 5 \end{Vmatrix}\right)\begin{Vmatrix} 2 & -1 \\ 0 & 3 \end{Vmatrix} = \begin{Vmatrix} 11 & 26 \\ 4 & 9 \end{Vmatrix}\begin{Vmatrix} 2 & -1 \\ 0 & 3 \end{Vmatrix} = \begin{Vmatrix} 22 & 67 \\ 8 & 23 \end{Vmatrix}$$

From the definition of conformable matrices, it is evident that a matrix is conformable to itself if and only if it is square Hence, a matrix A can be multiplied by itself if and only if it is square In such a case the product AA is referred to as the **square** of A and is denoted by the symbol A^2 Higher powers of A are defined in similar fashion, Theorem 2 guaranteeing that the definition

$$A^r = \underbrace{AA \cdots A}_{r \text{ factors}}$$

is unambiguous. With A^0 defined as the identity matrix I, it is obvious that, for any nonnegative integers r and s, the familiar laws of exponents

$$A^rA^s = A^{r+s} \qquad \text{and} \qquad (A^r)^s = A^{rs}$$

hold for matric multiplication

THEOREM 3

For suitably conformable matrices, multiplication is distributive over addition; i e., $A(B + C) = AB + AC$.

Theorems 2 and 3 are "obvious", that is, they assert properties which we know to be true for products in elementary algebra and which, by analogy, we would expect to be true in matric algebra. That these results must be proved and cannot be taken for granted is clear, however, from the next two theorems, which tell us that two other equally simple properties of ordinary algebraic multiplication do not hold for matric multiplication·

THEOREM 4

The product of two nonzero matrices may be a zero matrix, i e., the fact that $AB = O$ does not imply that either $A = O$ or $B = O$.

PROOF Clearly, to prove this theorem it is sufficient to exhibit two nonzero matrices whose product is a zero matrix, and we have, among infinitely many possibilities,

$$\begin{Vmatrix} 6 & 4 & 2 \\ 9 & 6 & 3 \\ -3 & -2 & -1 \end{Vmatrix} \begin{Vmatrix} 0 & 1 & -2 \\ -1 & 0 & 3 \\ 2 & -3 & 0 \end{Vmatrix} = \begin{Vmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix}$$

THEOREM 5

Even for matrices which are conformable in either order, multiplication is not commutative; i.e., in general, $AB \neq BA$

PROOF To prove this theorem it is sufficient to exhibit two matrices A and B such that $AB \neq BA$, and we have, specifically,

$$\begin{Vmatrix} 1 & 2 \\ 3 & 4 \end{Vmatrix} \cdot \begin{Vmatrix} 1 & 1 \\ 4 & 1 \end{Vmatrix} = \begin{Vmatrix} 9 & 3 \\ 19 & 7 \end{Vmatrix} \qquad \text{and} \qquad \begin{Vmatrix} 1 & 1 \\ 4 & 1 \end{Vmatrix} \begin{Vmatrix} 1 & 2 \\ 3 & 4 \end{Vmatrix} = \begin{Vmatrix} 4 & 6 \\ 7 & 12 \end{Vmatrix}$$

In two special cases, however, the multiplication of matrices is commutative. Though these are simple and obvious, they are of sufficient importance to be stated explicitly:

THEOREM 6

Both unit matrices and zero matrices commute with all suitably conformable matrices; more specifically, $AI = IA = A$ and $AO = OA = O$.

Since matric multiplication is not, in general, commutative, it is desirable to be able to describe concisely the order in which two conformable matrices are to be multiplied. This we shall do

by adopting the following terminology In the product AB we shall say that A **premultiplies** B, or B is **premultiplied** by A, and B **postmultiplies** A, or A is **postmultiplied** by B.

THEOREM 7

The transpose of the product of two conformable matrices is equal to the product of the transposed matrices taken in the other order; i.e., $(AB)^T = B^TA^T$

PROOF Let A be a (p,q) matrix and let B be a (q,r) matrix Then from the definition of the transpose of a matrix it follows that the element in the ith row and jth column of $(AB)^T$ is the element in the jth row and ith column of AB, namely,

$$\sum_{k=1}^{q} a_{jk}b_{ki} \tag{3}$$

On the other hand, the ith row of B^T is, by definition, the ith column of B; i.e., the ith row of B^T consists of the elements

$$b_{1i},\ b_{2i},\ \quad ,\ b_{qi}$$

Similarly, the jth column of A^T is, by definition, the jth row of A; i e , the jth column of A^T consists of the elements

$$a_{j1},\ a_{j2},\ \quad \cdot\ ,\ a_{jq}$$

Hence, the element in the ith row and jth column of B^TA^T is

$$\sum_{k=1}^{q} b_{ki}a_{jk} = \sum_{k=1}^{q} a_{jk}b_{ki}$$

Since this is the same as the expression (3) for the corresponding element in the matrix $(AB)^T$, the theorem is established.

COROLLARY 1

The transpose of the product of any number of conformable matrices is the product of the transposed matrices taken in the other order, i e.,

$$(A_1A_2 \quad \cdot A_n)^T = A_n{}^T \cdot \cdot \quad A_2{}^TA_1{}^T$$

The definition of a matrix in no way rules out the possibility that the elements of a matrix are themselves matrices In fact, it is often convenient to subdivide, or **partition**, a matrix into submatrices and then regard the original matrix as a new matrix having these submatrices as elements In particular, it is frequently helpful to regard an (m,n) matrix $A = \|a_{ij}\|$ as a row matrix $\|C_1\ C_2 \quad \cdot\ C_n\|$, whose elements are the respective column vectors of A, or as a column matrix

$$\begin{Vmatrix} R_1 \\ R_2 \\ \cdot \\ R_m \end{Vmatrix}$$

whose elements are the respective row vectors of A.

EXAMPLE 3

For instance, among numerous other possibilities, we can write

$$A = \left\| \begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \end{array} \right\| = \left\| \begin{matrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{matrix} \right\|$$

where $A_{11} = \left\| \begin{matrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{matrix} \right\| \qquad A_{12} = \left\| \begin{matrix} a_{14} \\ a_{24} \end{matrix} \right\| \qquad A_{21} = \| a_{31} \quad a_{32} \quad a_{33} \| \qquad A_{22} = \| a_{34} \|$

or equally well $A = \| A_{11} \quad A_{12} \quad A_{13} \quad A_{14} \|$

where now $A_{11} = \left\| \begin{matrix} a_{11} \\ a_{21} \\ a_{31} \end{matrix} \right\| \qquad A_{12} = \left\| \begin{matrix} a_{12} \\ a_{22} \\ a_{32} \end{matrix} \right\| \qquad A_{13} = \left\| \begin{matrix} a_{13} \\ a_{23} \\ a_{33} \end{matrix} \right\| \qquad A_{14} = \left\| \begin{matrix} a_{14} \\ a_{24} \\ a_{34} \end{matrix} \right\|$

In constructing the product of two matrices it is sometimes convenient to partition them before performing the multiplication. This can be done in many ways, but it is of course necessary that the given matrices be conformable and that the various submatrices which must be multiplied together also be conformable. This requirement imposes no restriction on the horizontal partitioning of the first matrix or on the vertical partitioning of the second matrix. It does require, however, that the columns of the first matrix be partitioned into groups such that the number of columns in each group is equal to the number of rows in the corresponding partition of the rows of the second matrix. Matrices for which this is the case are said to be **conformably partitioned.**

EXAMPLE 4

By direct multiplication we have

$$\left\| \begin{matrix} 1 & 1 & 1 \\ 2 & -1 & 0 \\ -1 & 0 & 2 \end{matrix} \right\| \left\| \begin{matrix} 1 & 2 & 3 & -1 \\ 3 & -1 & 1 & 0 \\ 0 & 0 & -2 & 1 \end{matrix} \right\| = \left\| \begin{matrix} 4 & 1 & 2 & 0 \\ -1 & 5 & 5 & -2 \\ -1 & -2 & -7 & 3 \end{matrix} \right\|$$

On the other hand we can write, among various other possibilities,

$$\left\| \begin{array}{cc|c} 1 & 1 & 1 \\ 2 & -1 & 0 \\ \hline -1 & 0 & 2 \end{array} \right\| = \left\| \begin{matrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{matrix} \right\| \qquad \text{and} \qquad \left\| \begin{array}{cccc} 1 & 2 & 3 & -1 \\ 3 & -1 & 1 & 0 \\ \hline 0 & 0 & -2 & 1 \end{array} \right\| = \left\| \begin{matrix} B_{11} \\ B_{21} \end{matrix} \right\|$$

and, from this point of view, the product of the two matrices is

$$\left\| \begin{matrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{matrix} \right\| \left\| \begin{matrix} B_{11} \\ B_{21} \end{matrix} \right\| = \left\| \begin{matrix} A_{11}B_{11} + A_{12}B_{21} \\ A_{21}B_{11} + A_{22}B_{21} \end{matrix} \right\|$$

or, performing the indicated multiplications and additions of the submatrices,

$$\left\| \begin{matrix} \left\| \begin{matrix} 4 & 1 & 4 & -1 \\ -1 & 5 & 5 & -2 \end{matrix} \right\| + \left\| \begin{matrix} 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 0 \end{matrix} \right\| \\ \| -1 \quad -2 \quad -3 \quad 1 \| + \| 0 \quad 0 \quad -4 \quad 2 \| \end{matrix} \right\| = \left\| \begin{matrix} 4 & 1 & 2 & 0 \\ -1 & 5 & 5 & -2 \\ -1 & -2 & -7 & 3 \end{matrix} \right\|$$

as before

Historically, the definition of the product of two matrices was introduced by the English mathematician Arthur Cayley (1821–1895) as a result of his investigations on linear transformations. By a **linear transformation** we mean a relation of the form

$$T_a\quad \begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\cdots\cdots\cdots\cdots \\ y_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{aligned}$$

connecting the variables $(x_1,x_2, \ldots ,x_n)$ and the variables $(y_1,y_2, \ldots ,y_n)$ in which the a_{ij} are constants independent of $(x_1, \ldots ,x_n)$ and $(y_1, \ldots ,y_n)$ If $n = 2$, we can think of T_a as a transformation of the cartesian plane which sends a point with coordinates (x_1,x_2) into a point with coordinates (y_1,y_2) Similarly, if $n = 3$, we can think of T_a as a transformation which sends a point with coordinates (x_1,x_2,x_3) into a point with coordinates (y_1,y_2,y_3). If $n > 3$, we can regard T_a as a transformation in a hyperspace of the appropriate number of dimensions, or we may think of it simply as a transformation of an n-component vector X into an n-component vector Y From the definition of matric multiplication and the equality of two matrices, it is clear that if we introduce the matrices

$$Y = \left\Vert \begin{matrix} y_1 \\ y_2 \\ \cdot \\ y_n \end{matrix} \right\Vert \qquad A = \left\Vert \begin{matrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{matrix} \right\Vert \qquad X = \left\Vert \begin{matrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{matrix} \right\Vert$$

then the transformation T_a can be written in the form

(4) $$T_a \qquad Y = AX$$

The matrix A in Eq (4) is usually referred to as the **matrix of the transformation** T_a It is thus apparent that matrices are intimately related to linear transformations and systems of linear equations

Suppose, now, that, in addition to the transformation T_a which transforms a vector X into a vector Y, we have a second transformation

(5) $$T_b\cdot \quad Z = BY \qquad Z = \left\Vert \begin{matrix} z_1 \\ z_2 \\ \\ z_n \end{matrix} \right\Vert \qquad B = \left\Vert \begin{matrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{matrix} \right\Vert$$

which transforms a vector Y into a vector Z If T_a is applied to a vector X and then T_b is applied to the resulting vector Y, the net result is to transform the vector X into the vector Z, and it is a matter of some interest to find the equations of the equivalent transformation connecting X directly with Z. This can easily be done, of course, simply by eliminating the variables $(y_1,y_2, \ldots ,y_n)$ between the equations of T_a and T_b To do this we observe that the equations of T_a and T_b can be written respectively

$$T_a. \quad y_k = \sum_{j=1}^{n} a_{kj}x_j \qquad k = 1, 2, \ldots, n$$

$$T_b. \quad z_i = \sum_{k=1}^{n} b_{ik}y_k \qquad i = 1, 2, \ldots, n$$

Hence, eliminating the y's by substituting for y_k in the equations of T_b, we have

$$z_i = \sum_{k=1}^{n} b_{ik}\left(\sum_{j=1}^{n} a_{kj}x_j\right) = \sum_{j=1}^{n}\left(\sum_{k=1}^{n} b_{ik}a_{kj}\right)x_j \qquad i = 1, 2, \ldots, n$$

Thus the coefficient of x_j in the equation for z_i is

$$\sum_{k=1}^{n} b_{ik}a_{kj}$$

which is precisely the element c_{ij} in the product BA. In other words, the matrix form of the single transformation T_{ba} equivalent to following T_a by T_b, may be found simply by eliminating Y between Eq. (4) and Eq (5)

$$Z = BY = B(AX) = (BA)X$$

Thus we have established the following important result·

THEOREM 8

The result of following a linear transformation T_a· $Y = AX$ with the transformation T_b $Z = BY$ is the single transformation T_{ba}· $Z = BAX$, whose matrix is the product BA of the matrices of T_b and T_a

As a further illustration of the importance of matric multiplication, we return to the idea of transition probabilities which we mentioned earlier in this section Suppose that a system S can exist in any of n states $S_1, S_2, \ldots, S_n$, and that by some random process the system may pass directly from the ith state to the jth state with probability $p_{ij}^{(1)}$· $(i,j = 1, 2, \ldots, n)$ Naturally, the system may also pass from the ith state to the jth state by first passing to some intermediate state, say the kth, and then passing from the kth state to the jth; and the calculation of these two-step transition probabilities is a matter of some importance. Now the probability that the system passes from the ith state to the jth state via the kth state is the product of the probability $p_{ik}^{(1)}$ that it passes in one step from S_i to S_k and the probability $p_{kj}^{(1)}$ that it then passes in one step from S_k to S_j Furthermore, since in any two-step transition from S_i to S_j the system must pass through *some* intermediate state (including, of course, S_i and S_j themselves) the probability $p_{ij}^{(2)}$ that the system passes in exactly two steps from S_i to S_j is

$$p_{ij}^{(2)} = \sum_{k=1}^{n} p_{ik}^{(1)}p_{kj}^{(1)}$$

But the sum on the right in the last formula is precisely the element in the ith row and jth column of the square of the matrix of one-step transition probabilities, $P = \|p_{ij}^{(1)}\|$. In other words, *the matrix of two-step transition probabilities for any system S is the square of the matrix of one-step transition probabilities for S.* A similar argument shows that the matrix of three-step transition probabilities for S is the cube of the matrix of one-step transition probabilities, and so on

EXAMPLE 5

Let S be the system consisting of two players A and B who begin with two dollars apiece and match coins until one or the other has no more money If the states of the system are defined by the number of dollars in A's possession, specifically, if the system is in the state S_{i+1} whenever A has i dollars ($i = 0, 1, 2, 3, 4$), find the matrix of one-step transition probabilities. Then, by raising this matrix to the second, third, and fourth powers, find the matrices containing the two-, three-, and four-step transition probabilities for S What is the probability that A will be ruined in at most four turns? What is the probability that A will be ruined in exactly four turns?

Clearly unless A or B is bankrupt, A must either win a dollar or lose a dollar on each turn, and the probability of each of these events is $\frac{1}{2}$ Hence, if A has i dollars ($i = 1, 2, 3$), that is, if the system is in the state S_{i+1} ($i = 1, 2, 3$), the probability of a one-step transition to S_i is $\frac{1}{2}$, the probability of a one-step transition to S_{i+2} is $\frac{1}{2}$, and the probability of a one-step transition to any other state is zero On the other hand, if the system is in the state S_1, that is, if A is bankrupt, the system remains in that state, so the probability of a one-step transition from S_1 to S_1 is 1, and the probability of any other transition from S_1 is 0 Similarly, if the system is in the state S_5, that is if B is bankrupt, the system remains in that state, hence the probability of a one-step transition from S_5 to S_5 is 1, and the probability of any other transition is 0 Thus the matrix of one-step transition probabilities is

$$P = \left\| \begin{matrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \end{matrix} \right\|$$

By multiplying this matrix by itself we find at once that the matrix of two-step transition probabilities is

$$P^2 = \left\| \begin{matrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \end{matrix} \right\|$$

Similarly, by computing P^3 and P^4, we find the matrices of three-step and four-step transition probabilities, respectively, to be

$$P^3 = \left\| \begin{matrix} 1 & 0 & 0 & 0 & 0 \\ \frac{5}{8} & 0 & \frac{1}{4} & 0 & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{8} & 0 & \frac{1}{4} & 0 & \frac{5}{8} \\ 0 & 0 & 0 & 0 & 1 \end{matrix} \right\| \quad \text{and} \quad P^4 = \left\| \begin{matrix} 1 & 0 & 0 & 0 & 0 \\ \frac{5}{8} & \frac{1}{8} & 0 & \frac{1}{8} & \frac{1}{8} \\ \frac{3}{8} & 0 & \frac{1}{4} & 0 & \frac{3}{8} \\ \frac{1}{8} & \frac{1}{8} & 0 & \frac{1}{8} & \frac{5}{8} \\ 0 & 0 & 0 & 0 & 1 \end{matrix} \right\|$$

The probability that A is ruined in at most four turns is simply the probability of a four-step transition from S_3 to S_1, namely, $p_{31}^{(4)} = \frac{3}{8}$, since among such transitions are included those in which the system reaches S_1 in less than four steps and then remains there. The probability that A is ruined in four turns and not before is the probability that S reaches S_1 in four steps but does not reach it in three steps or less, namely, $p_{31}^{(4)} - p_{31}^{(3)} = \frac{3}{8} - \frac{1}{4} = \frac{1}{8}$.

EXERCISES

1 Multiply the matrices

$$\left\|\begin{array}{cc:c} 1 & 2 & -1 \\ \hdashline 3 & 0 & 2 \end{array}\right\| \quad \text{and} \quad \left\|\begin{array}{cc} 3 & 1 \\ 1 & 3 \\ \hdashline 2 & 0 \end{array}\right\|$$

using the indicated partitioning Check by multiplying without regard to the partitioning

2 Evaluate the matric polynomial $X^3 - 4X^2 - X + 4I$ for each of the following matrices

a $\left\|\begin{matrix} 1 & -1 \\ 2 & 0 \end{matrix}\right\|$ **b** $\left\|\begin{matrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 1 \end{matrix}\right\|$ **c** $\left\|\begin{matrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{matrix}\right\|$

3 Verify that $(X - 3I)(X - 2I) = (X - 2I)(X - 3I) = X^2 - 5X + 6I$ for

$$X = \left\|\begin{matrix} 1 & 2 \\ 2 & -1 \end{matrix}\right\| \quad \text{and} \quad X = \left\|\begin{matrix} 1 & 2 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{matrix}\right\|$$

Do you think that this relation is an identity for all square matrices X?

4 Show that $\left\|\begin{matrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{matrix}\right\|^n = \left\|\begin{matrix} \cos n\theta & \sin n\theta \\ -\sin n\theta & \cos n\theta \end{matrix}\right\|$

5 Show that $\left\|\begin{matrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{matrix}\right\|^n = \left\|\begin{matrix} \cosh n\theta & \sinh n\theta \\ \sinh n\theta & \cosh n\theta \end{matrix}\right\|$

6 Show that, if a matrix A commutes with a diagonal matrix D whose diagonal elements are all distinct, then A is a diagonal matrix Is A necessarily diagonal if the diagonal elements of D are not all distinct?

7 If A and B are conformable matrices, show that the ith row vector in the product AB is A_iB where A_i is the ith row vector of A What is the jth column vector in the product AB?

8 If $A = \left\|\begin{matrix} 1 & 2 & -1 \\ 2 & 1 & 3 \\ 4 & 5 & 1 \end{matrix}\right\|$, find a nonzero 3×3 matrix X such that AX is a zero matrix Is $XA = O$? Is X unique?

9 If A and B are symmetric matrices of the same order, prove that the product AB is symmetric if and only if $AB = BA$

10 If A and B are two square matrices which commute and if r and s are positive integers, prove that A^r and B^s also commute

11 Prove Theorem 3

12 Prove Corollary 1, Theorem 7

13 Prove that $(A + B)^T = A^T + B^T$

14 If K is a diagonal matrix whose diagonal elements are all equal to k, prove that the product of K and any conformable matrix A is equal to the product of A and the scalar k Because of this property the matrix K is often referred to as a **scalar matrix.**

15 By definition, the transpose of the matrix $A = \|A_{ij}\|$ is the matrix $A^T = \|a_{ji}\|$ Is this formula correct if the a_{ij}'s are submatrices of A?

16 By the **derivative of a matrix** A we mean the matrix whose elements are the derivatives of the elements of A Assuming that the elements of the matrices A and B are differentiable functions of x, use this definition to show that $d(AB)/dx = (dA/dx)B + A(dB/dx)$ Is $dA^2/dx = 2A(dA/dx)$?

17 Show that in any matrix of transition probabilities, the sum of the elements in any row is 1

18 Consider the system S consisting of four boxes B_1, B_2, B_3, and B_4 and a single ball, and let the system be in the state S_i ($i = 1, 2, 3, 4$) if the ball is in the box B_i. Transitions from one state to another take place in the following manner A die is thrown and, if a 1, 2, or 3 turns up, the ball is taken from whichever box it may have been in and placed in the box bearing the number showing on the die If 4, 5, or 6 turns up, the ball is taken from wherever it may have been and placed in the box B_4 Find the matrix of one-, two-, and three-step transition probabilities for the system

19 Consider the system S consisting of three boxes B_1, B_2, and B_3 and a single ball, and let the system be in the state S_i ($i = 1, 2, 3$) if the ball is in the box B_i Transitions from one state to another take place in the following manner Three coins are tossed If no heads turn up, the ball is not moved, but, if one or more heads turn up, the ball is taken from whichever box it may have been in and placed in the box corresponding to the number of heads showing Find the matrix of one-, two-, three-, and four-step transition probabilities for the system

20 Show that, in computing the product of a (p,q) matrix and a (q,r) matrix, pqr multiplications and $pr(q - 1)$ additions must be performed If a (p,q) matrix and a (q,r) matrix are conformably partitioned into four submatrices by one partition of their rows and one partition of their columns, prove that the same number of multiplications and the same number of additions are required in multiplying the two matrices whether this is done in the original or in the partitioned form

10.3

Adjoints and inverses

It is a familiar fact of elementary algebra that any quantity Q not equal to zero has a reciprocal

$$Q^{-1} = \frac{1}{Q}$$

with the property that

$$QQ^{-1} = Q^{-1}Q = 1$$

The familiar process of division, which we sometimes inaccurately regard as being essentially independent of multiplication, is nothing but multiplication involving the reciprocal, or multiplicative inverse, of the divisor as one factor. In matric algebra, although we do not define division as such, we can in an important class of cases define the reciprocal, or inverse, of a matrix. With reciprocals defined, multiplication then serves to accomplish all we might properly expect to do by division As usual our development begins with a number of definitions.

We have already defined the determinant of a square matrix as the determinant whose array of elements is identical with the array of the matrix itself Clearly, only square matrices have determinants. A square matrix whose determinant is different from zero is said to be **nonsingular.** A square matrix whose determinant is equal to zero is said to be **singular.** Using these notions, we can now give formal definitions of the important concepts of the **adjoint** and the **inverse** of a matrix.

DEFINITION 1

If $A = \|a_{ij}\|$ is a square matrix and if A_{ij} is the cofactor of a_{ij} in the determinant of A, then the matrix

$$\|A_{ji}\| = \|A_{ij}\|^T = \text{transpose of } \|A_{ij}\|$$

is called the adjoint of the matrix A.

The adjoint of a square matrix A is sometimes indicated by the notation **adj** A.

DEFINITION 2

The reciprocal, or inverse, A^{-1} of a nonsingular matrix $A = \|a_{ij}\|$ is the adjoint of A divided by the determinant of A; i e ,

$$A^{-1} = \frac{\|A_{ij}\|^T}{|A|} = \frac{\|A_{ji}\|}{|A|}$$

Clearly, although every square matrix has an adjoint, only nonsingular matrices have inverses.

The fundamental importance of the reciprocal, or inverse, of a matrix is apparent from the following theorem

THEOREM 1

The product of a nonsingular matrix A and its reciprocal in either order is a unit matrix; i e , $A^{-1}A = AA^{-1} = I$

PROOF Let $A = \|a_{ij}\|$ be a nonsingular matrix and consider first the product

$$A^{-1}A = \frac{1}{|A|}\begin{Vmatrix} A_{11} & A_{21} & \cdot & A_{n1} \\ A_{12} & A_{22} & \cdot & A_{n2} \\ & & & \cdot \\ A_{1n} & A_{2n} & & A_{nn} \end{Vmatrix} \begin{Vmatrix} a_{11} & a_{12} & & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \cdot & & & \\ a_{n1} & a_{n2} & \cdot & a_{nn} \end{Vmatrix}$$

Clearly, from the definition of matric multiplication, the element in the ith row and jth column of the product of the two matrices on the right is the scalar product

$$\sum_{k=1}^{n} A_{ki}a_{kj}$$

Moreover, from Corollary 1, Theorem 11, Sec 10 1, this sum is equal to $|A|$ if $i = j$ and is equal to zero if $i \neq j$ Hence,

$$A^{-1}A = \frac{1}{|A|}\begin{Vmatrix} |A| & 0 & & 0 \\ 0 & |A| & & 0 \\ & & & \\ 0 & 0 & & |A| \end{Vmatrix} = \begin{Vmatrix} 1 & 0 & & 0 \\ 0 & 1 & & 0 \\ & & \cdot & \cdot\ \cdot \\ 0 & 0 & & 1 \end{Vmatrix} = I \qquad \text{as asserted}$$

The proof that $AA^{-1} = I$ follows in exactly the same fashion

COROLLARY 1

For any square matrix, $(\text{adj } A)A = A(\text{adj } A) = |A|I$.

COROLLARY 2

If A is a nonsingular matrix, then A^{-1} is also nonsingular and $|A^{-1}| = 1/|A|$

EXAMPLE 1

If

$$A = \begin{Vmatrix} 1 & 2 & 4 \\ -1 & 0 & 3 \\ 3 & 1 & -2 \end{Vmatrix}$$

then the determinant of A is

$$\begin{vmatrix} 1 & 2 & 4 \\ -1 & 0 & 3 \\ 3 & 1 & -2 \end{vmatrix} = 7$$

The adjoint of A is the transpose of

$$\begin{Vmatrix} \begin{vmatrix} 0 & 3 \\ 1 & -2 \end{vmatrix} & -\begin{vmatrix} -1 & 3 \\ 3 & -2 \end{vmatrix} & \begin{vmatrix} -1 & 0 \\ 3 & 1 \end{vmatrix} \\ -\begin{vmatrix} 2 & 4 \\ 1 & -2 \end{vmatrix} & \begin{vmatrix} 1 & 4 \\ 3 & -2 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} \\ \begin{vmatrix} 2 & 4 \\ 0 & 3 \end{vmatrix} & -\begin{vmatrix} 1 & 4 \\ -1 & 3 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ -1 & 0 \end{vmatrix} \end{Vmatrix} \quad \text{that is} \quad \begin{Vmatrix} -3 & 8 & 6 \\ 7 & -14 & -7 \\ -1 & 5 & 2 \end{Vmatrix}$$

The inverse of A is, therefore,

$$A^{-1} = \tfrac{1}{7}\begin{Vmatrix} -3 & 8 & 6 \\ 7 & -14 & -7 \\ -1 & 5 & 2 \end{Vmatrix}$$

and

$$A^{-1}A = \tfrac{1}{7}\begin{Vmatrix} -3 & 8 & 6 \\ 7 & -14 & -7 \\ -1 & 5 & 2 \end{Vmatrix}\begin{Vmatrix} 1 & 2 & 4 \\ -1 & 0 & 3 \\ 3 & 1 & -2 \end{Vmatrix} = \tfrac{1}{7}\begin{Vmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 7 \end{Vmatrix} = \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}$$

From Theorem 1 it is clear that, if A is a nonsingular matrix, then each of the equations $AX = I$ and $XA = I$ has $X = A^{-1}$ as *one* solution Actually, we have the following stronger result

THEOREM 2

If A is a nonsingular matrix, then $X = A^{-1}$ is the unique solution of each of the equations $AX = I$ and $XA = I$.

PROOF Consider first the equation $AX = I$ If both X_1 and X_2 satisfy this equation, then $AX_1 = AX_2$, since each is equal to I Moreover, since A is nonsingular, it follows that A^{-1} exists Hence, premultiplying by A^{-1}, we have

$$A^{-1}AX_1 = A^{-1}AX_2$$

$$IX_1 = IX_2$$

$$X_1 = X_2$$

Thus the equation $AX = I$ has in fact just one solution, and from Theorem 1 it follows that this solution is $X = A^{-1}$ A similar argument shows that $X = A^{-1}$ is also the unique solution of the equation $XA = I$, as asserted.

COROLLARY 1

If A is a nonsingular (n,n) matrix, if B is an (n,m) matrix, and if C is an (m,n) matrix, the equation $AX = B$ has the unique solution $X = A^{-1}B$, and the equation $XA = C$ has the unique solution $X = CA^{-1}$.

Various other important theorems follow easily now that the uniqueness of the solution of $AX = I$ for any nonsingular matrix A has been established. In particular, we have the following results:

THEOREM 3

If A is a nonsingular matrix, then $(A^{-1})^{-1} = A$.

PROOF By Theorem 2, $(A^{-1})^{-1}$ is the unique solution of the equation $A^{-1}X = I$. However, it is obvious by inspection that $X = A$ satisfies this equation Hence $(A^{-1})^{-1} = A$, as asserted

THEOREM 4

If A and B are nonsingular (n,n) matrices, then $(AB)^{-1} = B^{-1}A^{-1}$

PROOF By Theorem 2, $(AB)^{-1}$ is the unique solution of the equation $(AB)X = I$. However, it is clear that $X = B^{-1}A^{-1}$ satisfies this equation, since

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

Hence, $(AB)^{-1} = B^{-1}A^{-1}$, as asserted

COROLLARY 1

If $A, B, \ldots, K$ are nonsingular (n,n) matrices, then

$$(AB \cdots K)^{-1} = K^{-1} \cdots B^{-1}A^{-1}$$

With the inverse of a nonsingular matrix defined, it is now possible to define negative integral powers of any nonsingular matrix:

DEFINITION 3

If A is a nonsingular matrix and if r is a positive integer, then $A^{-r} = (A^{-1})^r$

Negative powers of singular matrices are not defined

For nonsingular matrices it is now possible to extend the familiar laws of exponents to negative as well as nonnegative integral powers

THEOREM 5

If A is a nonsingular matrix, then, for all integral values of r and s, $A^rA^s = A^{r+s}$ and $(A^r)^s = A^{rs}$.

If in the corollary of Theorem 2 we take B to be the column matrix

$$\left\| \begin{matrix} b_1 \\ b_2 \\ \cdot \\ b_n \end{matrix} \right\|$$

the matric equation $AX = B$ is equivalent to the system of nonhomogenous linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

Hence, it is clear from Corollary 1, Theorem 2, that the solution of this system exists and is uniquely given by

$$X = A^{-1}B$$

if A is nonsingular

Similarly, if we take B to be the matrix

$$Y = \begin{Vmatrix} y_1 \\ y_2 \\ \cdot \\ y_n \end{Vmatrix}$$

the matric equation $AX = B$ becomes the linear transformation

$$Y = AX$$

The corollary of Theorem 2 now assures us that if A is nonsingular the inverse of this transformation, that is, the transformation that carries us back from the vector Y to the vector X, is unique and has the equation

$$X = A^{-1}Y$$

As a physical illustration of the relation of a matrix and its inverse, let us consider the mass-spring system shown in Fig 10 1 and determine the forces which act on each of the masses as a result of arbitrary displacements x_1, x_2, and x_3 of the respective masses. The modulus of each spring is the indicated value of k; that is, the force required to stretch each spring a unit distance is the corresponding value of k Now, if the masses are displaced by the respective amounts x_1, x_2, and x_3, the increases in the length of the various springs are

k_1. x_1
k_{12}: $x_2 - x_1$
k_{13}: $x_3 - x_1$
k_{23}: $x_3 - x_2$
k_3: $-x_3$

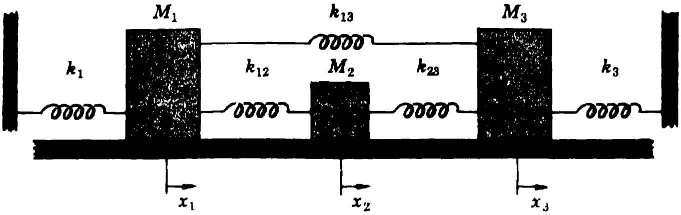


FIGURE 10.1 A typical system of spring-connected masses.

and the forces represented by these changes in length are

k_1: $k_1 x_1$
k_{12}: $k_{12}(x_2 - x_1)$
k_{13}: $k_{13}(x_3 - x_1)$
k_{23}: $k_{23}(x_3 - x_2)$
k_3: $k_3(-x_3)$

a positive force indicating that the spring is stretched and a negative force indicating that the spring is compressed. Hence, taking due account of the direction of the force applied to each mass by each spring attached to it, we find that the forces f_1, f_2, and f_3 which act on the respective masses are

(1)
$$\begin{aligned} f_1 &= -k_1 x_1 + k_{12}(x_2 - x_1) + k_{13}(x_3 - x_1) \\ f_2 &= -k_{12}(x_2 - x_1) + k_{23}(x_3 - x_2) \\ f_3 &= -k_{13}(x_3 - x_1) - k_{23}(x_3 - x_2) - k_3 x_3 \end{aligned}$$

or, collecting terms and rewriting in matric notation,

(2)
$$F = KX$$

where

$$F = \begin{Vmatrix} f_1 \\ f_2 \\ f_3 \end{Vmatrix} \qquad X = \begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix}$$

and

$$K = \begin{Vmatrix} -(k_1 + k_{12} + k_{13}) & k_{12} & k_{13} \\ k_{12} & -(k_{12} + k_{23}) & k_{23} \\ k_{13} & k_{23} & -(k_{13} + k_{23} + k_3) \end{Vmatrix}$$

Evaluating the first of Eqs. (1) for $x_1 = 1$, $x_2 = 0$, $x_3 = 0$, it is clear that $-(k_1 + k_{12} + k_{13})$ is the force applied to the first mass as a result of a unit displacement of that mass. A similar evaluation shows that, in general, the element in the ith row and jth column of K is the force applied to the ith mass as a result of a unit displacement of the jth mass. Because of this property the matrix K is usually referred to as the **stiffness matrix** of the system.

It can easily be verified that, for all positive values of the k's, the matrix K is nonsingular. Hence, for the physical system shown in Fig. 10.1, K^{-1} exists, and we can solve Eq. (2) for X, getting

$$X = K^{-1}F$$

Now, evaluating the right-hand side of this equation for a force vector with one component 1 and the rest 0, it follows that the element of K^{-1} in the ith row and jth column is the displacement produced in the ith mass as a result of a unit force applied to the jth mass. Because of this property, the matrix K^{-1} is usually referred to as the **elasticity matrix*** of the system. Our discussion has thus illustrated the important fact that *for any elastic system, the elasticity matrix is the inverse of the stiffness matrix, and vice versa.*

In the last section we defined a number of special matrices, and now, with the concept of the inverse of a matrix available, our list can be extended to include several additional important types. In the following table we bring together the types we have already defined as well as the new ones we are here introducing:

table 10.1

A
$\bar{A}$ = conjugate of A
A^T = transpose of A
$\bar{A}^T$ = associate of A
A^{-1} = inverse or reciprocal of A (A nonsingular)

Condition on A	Type
$A = \bar{A}$	Real
$A = -\bar{A}$	Imaginary
$A = A^T$	Symmetric
$A = -A^T$	Skew-symmetric
$A = \bar{A}^T$	Hermitian
$A = -\bar{A}^T$	Skew-hermitian
$A = (A^T)^{-1}$, i.e., $A^{-1} = A^T$ or $AA^T = I$	Orthogonal
$A = (\bar{A}^T)^{-1}$, i.e., $A^{-1} = \bar{A}^T$ or $A\bar{A}^T = I$	Unitary

Although we cannot go into the details of the matter, it is worth noting that orthogonal matrices derive their name from the fact that the matrix of a transformation which is a rotation of mutually perpendicular, or orthogonal, axes in two or three dimensions is always orthogonal. For instance, it is well known that in the cartesian plane the equations of a general rotation of axes are

$$\begin{aligned} x_1' &= x_1 \cos\alpha + x_2 \sin\alpha \\ x_2' &= -x_1 \sin\alpha + x_2 \cos\alpha \end{aligned} \qquad \text{or} \qquad X' = AX$$

$$\text{where} \qquad X' = \left\| \begin{matrix} x_1' \\ x_2' \end{matrix} \right\| \qquad A = \left\| \begin{matrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{matrix} \right\| \qquad X = \left\| \begin{matrix} x_1 \\ x_2 \end{matrix} \right\|$$

and it is easy to show that A is orthogonal by verifying that $AA^T = I$.

---

* The symmetry of the stiffness and elasticity matrices, which asserts, for instance, that the force acting on the ith mass as a result of a unit displacement of the jth mass is equal to the force acting on the jth mass as a result of a unit displacement of the ith mass, is an illustration of the famous reciprocity theorem of Maxwell-Rayleigh-Betti. This theorem is the counterpart in mechanics of what is known simply as the *reciprocity theorem* in electrical circuit analysis.

EXERCISES

1 Prove Corollary 1, Theorem 4.

2 Find the adjoint of each of the following matrices, and, when it exists, find the inverse

a $\begin{Vmatrix} 1 & 2 \\ 3 & 4 \end{Vmatrix}$ b $\begin{Vmatrix} 2 & -1 & 3 \\ 4 & 0 & -1 \\ 3 & 3 & 2 \end{Vmatrix}$ c $\begin{Vmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 3 & 2 & 1 \end{Vmatrix}$ d $\begin{Vmatrix} 2 & 3 & 1 \\ 1 & -1 & 2 \\ 1 & 9 & -4 \end{Vmatrix}$

3 a Under what conditions, if any, does $AB = AC$ imply $B = C$?
b If A is a nonsingular matrix, show that $AB = 0$ implies $B = 0$

4 If A is a nonsingular matrix which commutes with a matrix B, prove that A^{-1} commutes with B If B is also nonsingular, do A^{-1} and B^{-1} commute?

5 If D is a nonsingular diagonal matrix, prove that D^{-1} is also a diagonal matrix and that each element on the principal diagonal of D^{-1} is the reciprocal of the corresponding element in D Does a similar result hold for nonsingular triangular matrices?

6 If A is a singular matrix, prove that the product of A and its adjoint is a null matrix

7 Solve the system $\begin{aligned} x_1 - x_2 + 2x_3 &= 1 \\ 2x_1 \qquad - x_3 &= 2 \\ x_1 + x_2 + \; x_3 &= 3 \end{aligned}$ by multiplying both sides of the equivalent matric equation $AX = B$ by the inverse of the matrix of the coefficients, A^{-1}

8 If A is a nonsingular matrix, show that the determinant of the adjoint of A is equal to the $(n - 1)$st power of the determinant of A

9 If A is a nonsingular matrix, show that the adjoint of the adjoint of A is equal to A times the $(n - 2)$nd power of the determinant of A

10 Prove that the determinant of any orthogonal matrix is either 1 or -1 Is the converse true?

11 Prove that a real matrix is orthogonal if and only if its column vectors are unit vectors which are mutually orthogonal

12 Prove that, if the column vectors of a real matrix A are mutually orthogonal unit vectors, so are the row vectors of A

13 Show that, for all positive values of the k's, the matrix K in Eq (2) is nonsingular

14 Find the stiffness and elasticity matrices for the system shown in Fig 10 2

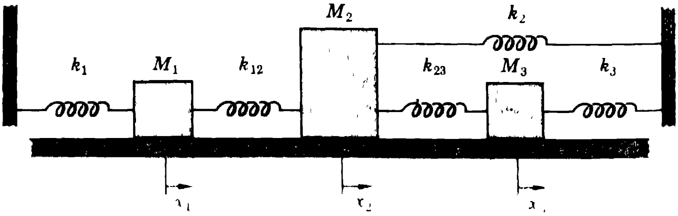


FIGURE 10 2

15 In mechanics it is shown that, if a cantilever beam bears a concentrated load P at a distance s from the fixed end, then the deflection y at a distance x from the fixed end is given by the formula*

$$y = \begin{cases} \dfrac{Px^2(x - 3s)}{6EI} & x \leqq s \\[2ex] \dfrac{Ps^2(s - 3x)}{6EI} & x \geqq s \end{cases}$$

---

* The theory required for the derivation of this result is summarized in Sec. 2.6.

where E and I are physical constants of the beam Using this formula, obtain the stiffness and elasticity matrices relating the forces and deflections at the positions $s = \frac{L}{3}, \frac{2L}{3}, L$ and $x = \frac{L}{3}, \frac{2L}{3}, L$. In what respect, if any, does this problem differ significantly from the example discussed in the text?

10.4

Rank and the equivalence of matrices

One of the most important characteristics of a matrix is its **rank**:

DEFINITION 1

The rank of a matrix A is the largest value of r for which there exists an (r,r) submatrix of A with nonvanishing determinant.

The rank of a matrix A, as we have just defined it, is sometimes referred to more specifically as the **determinant rank** of A. Clearly, as an immediate consequence of Theorem 1, Sec. 10 2, we have the following simple but useful result

THEOREM 1

If A and B are two (n,n) matrices of rank n, then both AB and BA are of rank n.

EXAMPLE 1

The matrix $\begin{Vmatrix} 1 & 2 & -1 & 3 \\ 3 & 4 & 0 & -1 \\ -1 & 0 & -2 & 7 \end{Vmatrix}$ is of rank 2, since each of the third-order submatrices

$$\begin{Vmatrix} 2 & -1 & 3 \\ 4 & 0 & -1 \\ 0 & -2 & 7 \end{Vmatrix} \quad \begin{Vmatrix} 1 & -1 & 3 \\ 3 & 0 & -1 \\ -1 & -2 & 7 \end{Vmatrix} \quad \begin{Vmatrix} 1 & 2 & 3 \\ 3 & 4 & -1 \\ -1 & 0 & 7 \end{Vmatrix} \quad \begin{Vmatrix} 1 & 2 & -1 \\ 3 & 4 & 0 \\ -1 & 0 & -2 \end{Vmatrix}$$

is singular while not all second-order submatrices are singular Specifically, the determinant of the 2×2 submatrix in the upper left-hand corner is different from zero.

In working with matrices it is frequently necessary to consider the effect of performing upon them certain simple manipulations known as **elementary transformations**:

DEFINITION 2

An elementary transformation of a matrix is any one of the following operations:

a The multiplication of each element of a row or a column by the same nonzero constant

b The interchange of two rows or of two columns

c The addition of any multiple of the elements of one row, or one column, to the corresponding elements of another row, or column, respectively

The most important property of elementary transformations is contained in the following theorem.

THEOREM 2

The rank of a matrix is not altered by any sequence of elementary transformations.

PROOF Let A be an arbitrary matrix, and let r be its rank. Then every minor of A of order greater than r is zero, and at least one minor of order r is different from zero To prove the theorem it is clearly sufficient to prove that no elementary transformation can change the rank of A.

Consider first an elementary transformation of type **a**. By Theorem 6, Sec 10.1, such a transformation cannot affect the vanishing or nonvanishing of any minor of A; hence it cannot alter the rank of A

A transformation of type **b**, on the other hand, may affect the vanishing or nonvanishing of the minor in some particular position in A. However, after a transformation of type **b** every submatrix in A exists *somewhere* in the resulting matrix, with at most two rows or two columns interchanged Hence, by Theorem 7, Sec. 10.1, if all minors of A of order greater than r are zero, the same thing will be true after the transformation is performed, and if at least one rth-order minor of A is different from zero, the same thing will be true after the transformation Thus no transformation of type **b** can alter the rank of A

Finally, no transformation of type **c** can alter the rank of A. For consider the transformation consisting of modifying the elements of the jth row by adding to them some multiple λ of the corresponding elements in the ith row (The case of column modification is handled by an identical argument.) Clearly, some of the $(r + 1)$st-order minors of A are unaffected by this transformation Specifically, any $(r + 1)$st-order minor involving neither the ith nor the jth rows, both the ith and the jth rows, or just the ith row will surely be unaffected, i.e., will be left equal to zero On the other hand, the value of an $(r + 1)$st-order minor involving the jth row but not the ith may conceivably be affected, since one of its rows (the jth) is modified by means of a row of elements (the ith) from outside the minor. However, by the addition theorem for determinants (Theorem 9, Sec 10 1) the modified determinant can be written in the form

$$|S_1| + \lambda|S_2|$$

where S_1 and S_2 are square submatrices of A of order $r + 1$ and hence singular, by hypothesis Thus no vanishing $(r + 1)$st-order minor of A can be transformed into one which is different from zero by a transformation of type **c**, that is, no transformation of type **c** can increase the rank of A On the other hand, no transformation of type **c** can decrease the rank of A, either For if this were the case, then the inverse transformation, that is, the transformation consisting of adding $-\lambda$ times the elements of the ith row to the corresponding elements of the jth row in the new matrix, would be a transformation of type **c** which restored the matrix to its original form and hence increased its rank to the original value r, and this we have just proved to be impossible Thus the rank of A, being neither increased nor decreased by a transformation of type **c**, is invariant under this transformation also, and our proof is complete.

DEFINITION 3

Two matrices A and B, one of which (and hence either of which) can be obtained from the other by a series of elementary transformations, are said to be equivalent

It is interesting and important to note that any elementary transformation involving the rows of a matrix A can be accomplished by premultiplying A by a unit matrix on whose rows the same elementary transformation has been performed, and any elementary transformation involving the columns of A can be accomplished by postmultiplying A by a unit matrix on whose columns the same elementary transformation has been performed. More specifically, we have the following theorems, whose proofs follow immediately from the definition of matric multiplication:

THEOREM 3

If A is an arbitrary (m,n) matrix and if M (N) is the matrix obtained from the identity matrix I_m (I_n) by multiplying the elements in the ith row (column) by λ, then the product MA (AN) is identical with A except for the ith row (column) which consists of the elements of the ith row (column) of A each multiplied by λ.

THEOREM 4

If A is an arbitrary (m,n) matrix and if M (N) is the matrix obtained from the identity matrix I_m (I_n) by interchanging its ith and jth rows (columns), then the product MA (AN) is identical with A except for the ith and jth rows (columns) which are interchanged

THEOREM 5

If A is an arbitrary (m,n) matrix and if M (N) is the matrix obtained from I_m (I_n) by adding to the elements of the jth row (column) λ times the corresponding elements in the ith row (column), then the product MA (AN) is identical with A except for the jth row (column) which consists of the elements of the jth row (column) of A plus λ times the corresponding elements in the ith row (column) of A

From the preceding theorems it is clear that a sequence of elementary transformations T_1, T_2, , T_k on the rows (columns) of an (m,n) matrix A can be accomplished by premultiplying (postmultiplying) A by a sequence of matrices M_1, M_2, . . , M_k (N_1,N_2, ,N_k) each obtained from the identity matrix I_m (I_n) by performing upon its rows (columns) the same elementary transformations The product $M_k \cdots M_2M_1$ (N_1N_2 N_k) of the matrices by which A is premultiplied (postmultiplied) can, of course, be expressed as a single matrix P (Q), necessarily of rank m (n) since it is the product of matrices each obtained from I_m (I_n) by elementary transformations and each therefore of rank m (n). We have thus established the following important theorem·

THEOREM 6

If A and B are equivalent matrices, then $B = PAQ$, where P and Q are nonsingular matrices

In view of the way the nonsingular matrices P and Q of the last theorem were obtained, it is interesting to inquire whether, conversely, any nonsingular matrix can be obtained from the corresponding identity matrix by elementary row transformations or elementary column transformations. The answer is Yes, as the following pair of theorems makes clear:

THEOREM 7

Any nonsingular (n,n) matrix can be reduced to the identity matrix I_n either by elementary row transformations or by elementary column transformations

PROOF Let $P = \|p_{ij}\|$ be an arbitrary nonsingular (n,n) matrix. Because P is nonsingular, at least one element in the first column must be different from zero; and it is no specialization to assume that $p_{11} \neq 0$, for if this is not the case the interchange of two rows will bring a nonzero element into the leading position. Since $p_{11} \neq 0$, the leading element in the first column may be reduced to 1 by multiplying each of the elements in the first row by $1/p_{11}$ Then by subtracting the appropriate multiple of the first row from each of the other rows we obtain the matrix

$$\begin{Vmatrix} 1 & r_{12} & r_{13} & \cdot & \cdot & r_{1n} \\ 0 & r_{22} & r_{23} & & \cdot & r_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & r_{n2} & r_{n3} & \cdot & \cdot & r_{nn} \end{Vmatrix},$$

Since the original matrix, and therefore the last one, is nonsingular, it follows that the submatrix

$$\begin{Vmatrix} r_{22} & \cdot & \cdot & r_{2n} \\ & \cdot & \cdot & \\ r_{n2} & \cdot & \cdot & r_{nn} \end{Vmatrix}$$

is nonsingular. Hence the same reduction can be applied to it, and thus, continuing the process sufficiently, we obtain an upper triangular matrix

$$\begin{Vmatrix} 1 & s_{12} & s_{13} & \cdot & \cdot & s_{1n} \\ 0 & 1 & s_{23} & \cdot & \cdot & s_{2n} \\ 0 & 0 & 1 & \cdot & \cdot & s_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 1 \end{Vmatrix}$$

Finally, working upward by row operations similar to those we have just employed, the elements above the diagonal can all be reduced to zero. This proves the assertion of the theorem for reductions involving only elementary row transformations. A similar argument shows that P can also be reduced to the identity by means of elementary column transformations. Thus the theorem is established.

THEOREM 8

Any nonsingular (n,n) matrix can be obtained from the identity matrix I_n by a sequence of elementary row transformations or a sequence of elementary column transformations.

PROOF If R is any nonsingular (n,n) matrix, we know from the last theorem that a sequence of elementary row transformations can be found which will reduce R to I_n. Then, by Theorems 3, 4, and 5, we know that there exist corresponding matrices $M_1, M_2, \ldots, M_k$ such that

$$I_n = M_k \cdots M_2M_1R$$

Postmultiplying this equation by R^{-1}, which of course exists since R is nonsingular, we have

$$(1) \qquad R^{-1} = M_k \cdots M_2M_1I_n$$

In other words, if we multiply I_n by the matrices corresponding to the elementary row operations by which we reduce R to I_n, the result is R^{-1}. Thus, if we have a nonsingular matrix P, which we wish to obtain from the corresponding identity matrix by a sequence of elementary row transformations, we need only take the matrix R in Eq. (1) to be P^{-1}, since $(P^{-1})^{-1} = P$; that is, to obtain P we need only apply to I the successive row operations by which P^{-1} is converted into I. A similar argument shows that an arbitrary nonsingular matrix P can also be obtained from the corresponding identity matrix by a sequence of elementary column transformations.

Incidentally, Eq. (1) provides a method for determining the inverse of a matrix R which is of some practical value, since the matrices $M_1, M_2, \ldots, M_k$ can easily be found from the straightforward process of reducing R to the identity matrix.

EXAMPLE 2

Find a sequence of elementary row transformations which will reduce the matrix

$$P = \begin{Vmatrix} 1 & 2 & 0 \\ 2 & 3 & -1 \\ -1 & -1 & 2 \end{Vmatrix}$$

to I_3, determine the matrices $M_1, M_2, \ldots,$ corresponding to these transformations, and use these results to compute the inverse of P.

By inspection it is clear that P can be reduced to I_3 by the following sequence of row operations:

$$T_1 \quad \text{Row } 2 - 2 \text{ Row } 1 \quad M_1 = \begin{Vmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix} \qquad M_1P = \begin{Vmatrix} 1 & 2 & 0 \\ 0 & -1 & -1 \\ -1 & -1 & 2 \end{Vmatrix}$$

$$T_2 \quad \text{Row } 3 + \text{Row } 1 \quad M_2 = \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{Vmatrix} \qquad M_2M_1P = \begin{Vmatrix} 1 & 2 & 0 \\ 0 & -1 & -1 \\ 0 & 1 & 2 \end{Vmatrix}$$

$$T_3 \quad - \text{Row } 2 \quad M_3 = \begin{Vmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{Vmatrix} \qquad M_3M_2M_1P = \begin{Vmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{Vmatrix}$$

$$T_4 \quad \text{Row } 3 - \text{Row } 2 \quad M_4 = \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{Vmatrix} \qquad M_4M_3M_2M_1P = \begin{Vmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{Vmatrix}$$

$$T_5 \quad \text{Row } 2 - \text{Row } 3 \quad M_5 = \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{Vmatrix} \qquad M_5M_4M_3M_2M_1P = \begin{Vmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}$$

$$T_6. \quad \text{Row } 1 - 2 \cdot \text{Row } 2 \quad M_6 = \begin{Vmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix} \qquad M_6M_5M_4M_3M_2M_1P = \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix},$$

The inverse P^{-1} of the given matrix can now be found either by using Eq. (1), namely,

$$P^{-1} = M_6M_5M_4M_3M_2M_1I_3$$

or simply by performing on I_3 the same sequence of row transformations used to reduce P to I_3

$$P^{-1} = T_6T_5T_4T_3T_2T_1I_3$$

The result, by either method, is

$$P^{-1} = \begin{Vmatrix} -5 & 4 & 2 \\ 3 & -2 & -1 \\ -1 & 1 & 1 \end{Vmatrix}$$

With Theorem 8, we can now prove the converse of Theorem 6:

THEOREM 9

If $B = PAQ$, where P and Q are nonsingular matrices, then A and B are equivalent.

PROOF By Theorem 8, P can be obtained from the corresponding identity matrix by a sequence of elementary row transformations, and Q can be obtained from the corresponding identity matrix by a sequence of elementary column transformations Thus, as in the proof of Theorem 8, there is a set of matrices $M_1, M_2, \ldots, M_k$, each representing some elementary row operation, such that

$$P = M_k \cdots M_2M_1$$

and a set of matrices $N_1, N_2, \ldots, N_l$, each representing some elementary column operation, such that

$$Q = N_1N_2 \cdots N_l$$

Hence, $$B = PAQ = (M_k \cdots M_2M_1)A(N_1N_2 \cdots N_l)$$

which proves that B is obtained from A by elementary row and column transformations and, hence, is equivalent to A, as asserted

By a proof almost identical with the proof of Theorem 7, we can prove the following theorem:

THEOREM 10

Any (m,n) matrix of rank r can be reduced by elementary transformations, which in general will involve both rows and columns, to an (m,n) matrix in which $a_{ii} = 1$ $(i = 1, 2, \ldots, r)$ and all other elements are zero

From Theorem 2 it is clear that equivalent matrices have the same rank. In view of Theorem 10, it is clear that, given two (m,n) matrices of the same rank, each can be reduced by elementary transformations to the same standard form, and, hence, each can be reduced to the other, via the standard form, by elementary transformations. Thus we have established the following important theorem:

THEOREM 11

Two (m,n) matrices are equivalent if and only if they have the same rank.

The equivalence relation

$$B = PAQ \qquad P,Q \text{ nonsingular}$$

is a very general one, and many applications involve special cases in which P and Q satisfy additional conditions. These can all be thought of as transformations of a matrix A into a matrix B, and the usual terminology reflects this point of view. The following table summarizes the various cases of particular interest:

table 10 2

If P,Q are arbitrary nonsingular matrices	$B = PAQ$	is an equivalence transformation and B is equivalent to A
If $P = Q^{-1}$	$B = Q^{-1}AQ$	is a similarity transformation and B is similar to A
If $P = Q^T$	$B = Q^TAQ$	is a congruence transformation and B is congruent to A
If $P = Q^T = Q^{-1}$	$B = Q^TAQ$ $= Q^{-1}AQ$	is an orthogonal transformation and B is orthogonally similar to A.
If $P = \bar{Q}^T = Q^{-1}$	$B = \bar{Q}^TAQ$ $= Q^{-1}AQ$	is a unitary transformation and B is unitarily similar to A

EXERCISES

1 If a matrix is of rank r, is it possible that for some value ρ, less than r, all minors of order ρ are equal to zero? Why?

2 Determine the rank of each of the following matrices

a $\begin{Vmatrix} 3 & -1 & 2 & 4 \\ 6 & 2 & -4 & -8 \end{Vmatrix}$ **b** $\begin{Vmatrix} 2 & -1 & 3 \\ 1 & -2 & 3 \\ 5 & 0 & 3 \end{Vmatrix}$

c $\begin{Vmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 0 & 5 & -10 \end{Vmatrix}$ **d** $\begin{Vmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \\ 5 & 6 & 7 & 8 \end{Vmatrix}$

3 Determine the rank of each of the following matrices as a function of λ

a $\begin{Vmatrix} 8(1-\lambda) & -2 & 0 \\ -2 & 3-2\lambda & -1 \\ 0 & -1 & 2(1-\lambda) \end{Vmatrix}$ **b** $\begin{Vmatrix} 1-\lambda & 1 & 1 \\ 1 & 3-\lambda & 3 \\ 2 & 1 & 4-\lambda \end{Vmatrix}$

c $\begin{Vmatrix} 5-\lambda & 4 & -2 \\ 4 & 5-\lambda & -2 \\ -2 & -2 & 3-2\lambda \end{Vmatrix}$

4 If A is an $(m,1)$ matrix and B is a $(1,n)$ matrix, show that the rank of the matrix AB is 1

5 Prove Theorem 10

6 Prove that the relation of equivalence has the following properties:

- **a** Every matrix is equivalent to itself
- **b** If A is equivalent to B, then B is equivalent to A.
- **c** If A is equivalent to B and B is equivalent to C, then A is equivalent to C.

Do the other relations listed in Table 10 2 have these properties?

7 **a** Work Example 2 using only elementary column transformations

b Work Example 2 if P is the matrix $\begin{Vmatrix} 2 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{Vmatrix}$.

8 Show that $A = \begin{Vmatrix} 0 & 1 & 0 \\ 1 & 2 & 1 \end{Vmatrix}$ and $B = \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{Vmatrix}$ are equivalent, and find nonsingular matrices P and Q, such that $B = PAQ$

9 Show that $A = \begin{Vmatrix} 0 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{Vmatrix}$ and $B = \begin{Vmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 2 & 1 & 1 \end{Vmatrix}$ are equivalent, and find nonsingular matrices P and Q such that $B = PAQ$

10 Show that $A = \begin{Vmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 4 & -1 & -2 \end{Vmatrix}$ and $B = \begin{Vmatrix} 2 & 1 & 0 \\ 1 & 1 & 2 \\ 0 & -1 & -4 \end{Vmatrix}$ are equivalent, and find nonsingular matrices P and Q such that $B = PAQ$

10.5

Systems of linear equations

Determinants and matrices find their most important application in the study of **linear dependence** and **independence** and in the closely related problem of the solution of systems of simultaneous linear equations

DEFINITION 1

The quantities $Q_1, Q_2, \ldots, Q_n$ are said to be linearly dependent if there exists a set of constants $c_1, c_2, \ldots, c_n$, at least one of which is different from zero, such that the equation

$$c_1Q_1 + c_2Q_2 + \cdots + c_nQ_n = 0$$

holds identically

DEFINITION 2

The quantities $Q_1, Q_2, \ldots, Q_n$ are said to be linearly independent if they are not linearly dependent; i e, if the only linear equation of the form

$$c_1Q_1 + c_2Q_2 + \cdots + c_nQ_n = 0$$

which they satisfy identically has

$$c_1 = c_2 = \cdots = c_n = 0$$

THEOREM 1

If the quantities $Q_1, Q_2, \ldots, Q_n$ are linearly dependent, then at least one (though not necessarily each one) of the quantities can be expressed as a linear combination of the remaining ones.

PROOF Since $Q_1, Q_2, \ldots, Q_n$ are linearly dependent, they necessarily satisfy a linear equation of the form $c_1Q_1 + c_2Q_2 + \cdots + c_nQ_n = 0$ in which at least one of the c's, say c_i, is different from zero This being the case, we can divide by c_i, getting

$$Q_i = -\frac{c_1}{c_i}Q_1 - \frac{c_2}{c_i}Q_2 - \cdots - \frac{c_n}{c_i}Q_n$$

which expresses Q_i as a linear combination of the remaining Q's, as asserted Since some, though not all, of the c's may be zero, it follows that we may not be able to solve for each of the Q's in this fashion

EXAMPLE 1

Show that the quantities 1, x, and x^2 are linearly independent

If 1, x, and x^2 are not linearly independent, they must satisfy identically some linear equation of the form

$$c_1(1) + c_2(x) + c_3(x^2) = 0$$

in which at least one of the c's is different from zero However, evaluating this identity for the particular values $x = -1, 0, 1$, we obtain the three equations

$$\begin{aligned} c_1 - c_2 + c_3 &= 0 \\ c_1 \qquad\quad &= 0 \\ c_1 + c_2 + c_3 &= 0 \end{aligned}$$

and, by inspection, the only solution of this system is $c_1 = c_2 = c_3 = 0$. Since this contradicts the assumption of linear dependence, the given quantities must be linearly independent, as asserted

EXAMPLE 2

Show that the vectors

$$V_1 = \begin{Vmatrix} 1 \\ 2 \\ 3 \end{Vmatrix} \qquad V_2 = \begin{Vmatrix} 2 \\ -1 \\ 3 \end{Vmatrix} \qquad V_3 = \begin{Vmatrix} 0 \\ 1 \\ -1 \end{Vmatrix} \qquad V_4 = \begin{Vmatrix} 4 \\ -1 \\ 5 \end{Vmatrix}$$

are linearly dependent

These vectors will be linearly dependent if and only if constants c_1, c_2, c_3, and c_4 exist such that

a At least one of them is different from zero

b $$c_1\begin{Vmatrix} 1 \\ 2 \\ 3 \end{Vmatrix} + c_2\begin{Vmatrix} 2 \\ -1 \\ 3 \end{Vmatrix} + c_3\begin{Vmatrix} 0 \\ 1 \\ -1 \end{Vmatrix} + c_4\begin{Vmatrix} 4 \\ -1 \\ 5 \end{Vmatrix} = \begin{Vmatrix} 0 \\ 0 \\ 0 \end{Vmatrix}.$$

Condition **b** is, of course, equivalent to the three scalar equations

$$\begin{aligned} c_1 + 2c_2 \qquad\quad + 4c_4 &= 0 \\ 2c_1 - c_2 + c_3 - c_4 &= 0 \\ 3c_1 + 3c_2 - c_3 + 5c_4 &= 0 \end{aligned}$$

and it is not difficult to verify that these are satisfied by the values

$$c_1 = 0 \qquad c_2 = 2\lambda \qquad c_3 = \lambda \qquad c_4 = -\lambda \qquad \lambda \text{ arbitrary}$$

and by no others. Hence the four vectors are linearly dependent and, in fact, are connected by the relation

$$0V_1 + 2V_2 + V_3 - V_4 = O$$

and (except for constant multiples of this) no others. From this it is obvious that V_2, V_3, and V_4 can each be expressed in terms of the remaining vectors of the set, but that V_1 cannot be so expressed.*

On the other hand, the vectors V_1, V_2, and V_3 are linearly independent, since an equation of the form $c_1V_1 + c_2V_2 + c_3V_3 = O$, that is,

$$c_1\begin{Vmatrix} 1 \\ 2 \\ 3 \end{Vmatrix} + c_2\begin{Vmatrix} 2 \\ -1 \\ 3 \end{Vmatrix} + c_3\begin{Vmatrix} 0 \\ 1 \\ -1 \end{Vmatrix} = \begin{Vmatrix} 0 \\ 0 \\ 0 \end{Vmatrix}$$

implies that

$$\begin{aligned} c_1 + 2c_2 \qquad &= 0 \\ 2c_1 - c_2 + c_3 &= 0 \\ 3c_1 + 3c_2 - c_3 &= 0 \end{aligned}$$

and by direct solution we find that the only values which satisfy this system of equations are $c_1 = c_2 = c_3 = 0$. Similarly we can verify that V_1, V_2, and V_4 are independent and that V_1, V_3, and V_4 are independent. However V_2, V_3, and V_4 are dependent, since, as we observed above, they satisfy the relation $2V_2 + V_3 - V_4 = O$.

As a simple application of the notion of linear independence, we have the following useful result.

THEOREM 2

If $V_1, V_2, \dots, V_m$ are m vectors each having $n \geqq m$ components and if for $i = 1, 2, \dots, m$ the first (the last) nonzero component of V_i is the ith [the $(n - m + i)$th], then $V_1, V_2, \dots, V_m$ are linearly independent.

PROOF By hypothesis, the given vectors are of the form

$$V_1 = \begin{Vmatrix} v_{11} \\ v_{21} \\ v_{31} \\ \\ v_{m1} \\ \cdot \\ v_{n1} \end{Vmatrix} \qquad V_2 = \begin{Vmatrix} 0 \\ v_{22} \\ v_{32} \\ \\ v_{m2} \\ \\ v_{n2} \end{Vmatrix} \qquad V_3 = \begin{Vmatrix} 0 \\ 0 \\ v_{33} \\ \\ v_{m3} \\ \\ v_{n3} \end{Vmatrix} \qquad V_m = \begin{Vmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ v_{mm} \\ \\ v_{nm} \end{Vmatrix}$$

where $v_{11}, v_{22}, v_{33}, \dots, v_{mm}$ are all different from zero. Now, since each vector has n components, the condition $c_1V_1 + c_2V_2 + \cdots + c_mV_m = O$ implies n scalar equations, the first m of which are

$$\begin{aligned} c_1v_{11} \qquad\qquad\qquad\qquad\qquad\qquad &= 0 \\ c_1v_{21} + c_2v_{22} \qquad\qquad\qquad\qquad &= 0 \\ c_1v_{31} + c_2v_{32} + c_3v_{33} \qquad\qquad &= 0 \\ \cdots\cdots\cdots\cdots\cdots & \\ c_1v_{m1} + c_2v_{m2} + c_3v_{m3} + \cdots + c_mv_{mm} &= 0 \end{aligned}$$

---

* Of course it is possible for a set of dependent quantities $Q_1, Q_2, \dots, Q_n$ to satisfy more than one independent linear equation. In problems where this occurs, it may well be that some of the equations can be solved for Q_i, say, while the others cannot. Naturally, if even one equation can be solved for Q_i, then Q_i can be expressed in terms of the other members of the set.

Hence, since $v_{11}, v_{22}, v_{33}, \ldots, v_{mm}$ are all different from zero, it follows that $c_1 = c_2 = c_3 = \cdots = c_m = 0$ Therefore the vectors $V_1, V_2, V_3, \ldots, V_m$ are linearly independent, as asserted A similar argument establishes the parenthetical assertion of the theorem.

From Examples 1 and 2, it is clear that questions concerning linear dependence and independence are closely related to the solution of systems of simultaneous linear equations, and to these we now turn our attention In the most general case we have a system of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots &\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{1}$$

where m, the number of equations, is not necessarily equal to n, the number of unknowns If at least one of the m quantities b_i is different from zero, the system is said to be **nonhomogeneous.** If $b_i = 0$ for all values of i, the system is said to be **homogeneous.** If we define the matrices

$$A = \begin{Vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{Vmatrix} \qquad X = \begin{Vmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{Vmatrix} \qquad B = \begin{Vmatrix} b_1 \\ b_2 \\ \cdot \\ b_m \end{Vmatrix}$$

the system can be written in the compact matric form

$$AX = B \tag{2}$$

In this form, the matrix A is known as the **coefficient matrix** of the system and the matrix

$$\|AB\|$$

obtained by adjoining the column matrix B to the coefficient matrix A is known as the **augmented matrix** of the system.

Before proceeding to the question of the existence and determination of solutions of (2), we shall first prove several important theorems about such solutions on the assumption that they exist:*

THEOREM 3

If X_1 and X_2 are two solution vectors of the homogeneous matric equation $AX = O$, then, for all values of the scalar constants c_1 and c_2, the vector $c_1X_1 + c_2X_2$ is also a solution of $AX = O$

PROOF By direct substitution we have

$$\begin{aligned} A(c_1X_1 + c_2X_2) &= A(c_1X_1) + A(c_2X_2) \\ &= c_1(AX_1) + c_2(AX_2) \\ &= c_1 \cdot O + c_2 \cdot O \\ &= O \end{aligned}$$

---

* It is interesting to note the striking resemblance between the next three theorems and Theorems 1, 2, 3 of Sec. 2.1 and Theorems 1, 2, 3 of Sec 4 5.

where the coefficients of c_1 and c_2 vanish because, by hypothesis, both X_1 and X_2 are solutions of $AX = O$. Hence, $c_1X_1 + c_2X_2$ also satisfies $AX = O$, as asserted

THEOREM 4

If k is the maximum number of linearly independent solution vectors of the system $AX = O$ and if $X_1, X_2, \ldots, X_k$ are k particular linearly independent solution vectors, then any solution vector of $AX = O$ can be expressed in the form

$$c_1X_1 + c_2X_2 + \cdots + c_kX_k$$

where the c's are scalar constants.

PROOF Let k be the maximum number of linearly independent solution vectors of the equation $AX = O$, let $X_1, X_2, \ldots, X_k$ be a particular set of k linearly independent solution vectors, and let X_{k+1} be *any* solution vector If X_{k+1} is one of the vectors in the set $\{X_1,X_2, \ldots ,X_k\}$ the assertion of the theorem is obviously true If X_{k+1} is not a member of the set $\{X_1,X_2, \ldots ,X_k\}$ then $X_1, X_2, \ldots, X_k, X_{k+1}$ cannot be linearly independent, since, by hypothesis, k is the maximum number of linearly independent solution vectors of $AX = O$ Hence, the X's must satisfy a linear equation of the form

$$c_1X_1 + c_2X_2 + \cdots + c_kX_k + c_{k+1}X_{k+1} = O \tag{3}$$

in which at least one c is different from zero In fact, $c_{k+1} \neq 0$, for otherwise Eq. (3) would reduce to

$$c_1X_1 + c_2X_2 + \cdots + c_kX_k = O$$

with at least one of the c's different from zero, contrary to the hypothesis that $X_1, X_2, \ldots, X_k$ are linearly independent But if $c_{k+1} \neq 0$, it is clearly possible to solve Eq (3) for X_{k+1} and express it in the form asserted by the theorem

Because of the property guaranteed by the last theorem, a general linear combination of the maximum number of linearly independent solution vectors of $AX = O$ is usually referred to as a **complete solution** of $AX = O$.

THEOREM 5

If X_p is a particular solution vector of the nonhomogeneous system $AX = B$ and if $c_1X_1 + c_2X_2 + \cdots + c_kX_k$ is a complete solution of the related homogeneous system $AX = O$, then any solution of the nonhomogeneous system can be written in the form

$$c_1X_1 + c_2X_2 + \cdots + c_kX_k + X_p$$

PROOF Let X_p be a particular solution vector of the nonhomogeneous equation $AX = B$ and let X_a be any solution vector of this equation. Then $AX_p = B$ and $AX_a = B$, and, subtracting these two equations, we have

$$AX_a - AX_p = O \qquad \text{or} \qquad A(X_a - X_p) = O$$

Now the last equation shows that $X_a - X_p$ is a solution of the homogeneous equation $AX = O$. Hence, by Theorem 4, it can be expressed in the form

$$X_a - X_p = c_1X_1 + c_2X_2 + \cdots + c_kX_k$$

Therefore, transposing,

$$X_a = c_1X_1 + c_2X_2 + \cdots + c_kX_k + X_p$$

Since X_a was *any* solution vector of the equation $AX = B$, the theorem is established.

We now turn our attention to the question of when solutions of the equation $AX = B$ will actually exist. The central result is contained in the following theorem.

THEOREM 6

A system of m simultaneous linear equations in n unknowns, $AX = B$, has a solution if and only if the coefficient matrix A and the augmented matrix $\|AB\|$ have the same rank r. When solutions exist, the maximum number of independent arbitrary constants in any general solution, i.e., the maximum number of linearly independent solution vectors of the related homogeneous system $AX = O$, is $n - r$.

PROOF We shall prove this theorem by applying to the given system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

a procedure, known as the **Gauss reduction,** which resembles closely the method by which we proved Theorem 7, Sec. 10.4. We begin by assuming that $a_{11} \neq 0$, which is no specialization, since at least one of the coefficients in the first equation must be different from zero, and, by renaming the unknowns, if necessary, it can be brought into the leading position. Next we divide the first equation by a_{11} and then multiply it in turn by $a_{21}, a_{31}, \ldots, a_{m1}$ and subtract it from the second, third, . . . , mth equation. This gives the equivalent* system

$$\begin{aligned} x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \cdots + \alpha_{1n}x_n &= \beta_1 \\ a'_{22}x_2 + a'_{23}x_3 + \cdots + a'_{2n}x_n &= b'_2 \\ a'_{32}x_2 + a'_{33}x_3 + \cdots + a'_{3n}x_n &= b'_3 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a'_{m2}x_2 + a'_{m3}x_3 + \cdots + a'_{mn}x_n &= b'_m \end{aligned}$$

where, explicitly,

$$\alpha_{1j} = \frac{a_{1j}}{a_{11}} \qquad \beta_1 = \frac{b_1}{a_{11}} \qquad a'_{ij} = a_{ij} - a_{i1}\frac{a_{1j}}{a_{11}} \qquad b'_i = b_i - a_{i1}\frac{b_1}{a_{11}}$$

Now we apply the same process to the last $m - 1$ equations, noting that, if $a'_{22} = 0$, a renaming of the last $n - 1$ unknowns with possibly a rearrangement of the last $m - 1$ equations will introduce a nonzero coefficient in place of a'_{22} unless all coefficients in the remaining equations are zero, which, of course, may be the case at some stage in the process The result of this second reduction is the system

---

* **Two equations or systems of equations are said to be equivalent if every solution of one is a solution of the other, and conversely.**

$$\begin{array}{rl}
x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \cdots + \alpha_{1n}x_n & = \beta_1 \\
x_2 + \alpha_{23}x_3 + \cdots + \alpha_{2n}x_n & = \beta_2 \\
a''_{33}x_3 + \cdots + a''_{3n}x_n & = b''_3 \\
a''_{43}x_3 + \cdots + a''_{4n}x_n & = b''_4 \\
\cdots \qquad \cdots \qquad \cdots & \cdots \\
a''_{m3}x_3 + \cdots + a''_{mn}x_n & = b''_m
\end{array}$$

We now continue in exactly the same fashion until the process terminates If $m < n$, this may happen because after m applications there are no more equations to which to apply it·

$$\begin{array}{rl}
x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \cdots + \alpha_{1m}x_m + \alpha_{1,m+1}x_{m+1} + \cdots + \alpha_{1n}x_n & = \beta_1 \\
x_2 + \alpha_{23}x_3 + \cdot + \alpha_{2m}x_m + \alpha_{2,m+1}x_{m+1} + \cdot + \alpha_{2n}x_n & = \beta_2 \\
x_3 + \quad + \alpha_{3m}x_m + \alpha_{3,m+1}x_{m+1} + \quad + \alpha_{3n}x_n & = \beta_3 \\
\cdots \qquad \cdots \qquad \cdots & \cdot \\
x_m + \alpha_{m,m+1}x_{m+1} + \quad + \alpha_{mn}x_n & = \beta_m
\end{array}$$

On the other hand, regardless of the relative size of m and n, the process may terminate because before we have made m reductions, say after only k $(<m)$ reductions, all coefficients in the left member of each of the remaining $m - k$ equations are zero

$$\begin{array}{rlccll}
x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + & + \alpha_{1k}x_k + \cdot & + & + \cdot & + \alpha_{1n}x_n & = \beta_1 \\
x_2 + \alpha_{23}x_3 + & + \alpha_{2k}x_k + \cdot & + & + & + \alpha_{2n}x_n & = \beta_2 \\
x_3 + \cdot & + \alpha_{3k}x_k + & + & \cdot + & + \alpha_{3n}x_n & = \beta_3 \\
 & + \cdot \quad + & + & \cdot + & + & \cdot \quad \cdot \\
 & x_k + \cdot \cdot & + & + \cdot & + \alpha_{kn}x_n & = \beta_k \\
 & & 0 + & + \cdot & + \quad 0 & = \beta_{k+1} \\
 & & \cdot \cdot & + & + & \cdot \\
 & & & 0 + \cdot & + \quad 0 & = \beta_m
\end{array}$$

In the first case, if we transpose all terms containing $x_{m+1}, x_{m+2}, \qquad , x_n$, we have a system of equations from which $x_m, x_{m-1}, x_{m-2}, \qquad , x_2, x_1$ can successively be found in terms of $x_{m+1}, x_{m+2}, \qquad , x_n$, which can be given arbitrary values. In the second case it may be that

$$\beta_{k+1} = \beta_{k+2} = \qquad = \beta_m = 0$$

so that we have essentially the case we have just discussed, except that now it is $x_1, x_2, \ldots \quad , x_k$ which are expressed in terms of the remaining unknowns x_{k+1}, $x_{k+2}, \ldots \quad , x_n$, which can be given arbitrary values. If, however, one or more of the β's after β_k is different from zero, then we have a contradiction, and the original system has no solution, or, in other words, is **inconsistent.**

Now consider the coefficient matrix and the augmented matrix of the reduced system Since the augmented matrix contains the coefficient matrix, it is clear that its rank must be at least as great as the rank of the coefficient matrix Moreover, in the solvable cases it is evident from the definition of rank that the rank of the augmented matrix cannot exceed the rank of the coefficient matrix and, hence, must be equal to it. Likewise, it is obvious that in the inconsistent case the rank of the augmented matrix is actually greater than the rank of the coefficient matrix. Finally, we observe that the ranks of the coefficient matrix and the augmented matrix for the reduced system are equal to the ranks of the respective matrices of the original system, since each step in the Gauss reduction, namely, rearranging the columns of unknowns, rearranging the equations, multiplying and dividing the equations by nonzero constants, and subtracting multiples of one equation from other equations, is an elementary transformation, which, by Theorem 2, Sec. 10.4, cannot change the rank of either matrix. Thus we have established the first assertion of the theorem.

Let us now return to the reduced system in the solvable case. If r is the common value of the rank of the coefficient matrix and the augmented matrix, the reduced system can be written

$$\begin{array}{rcl}
x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \cdots + \alpha_{1r}x_r &=& -\alpha_{1,r+1}x_{r+1} - \cdots - \alpha_{1n}x_n + \beta_1 \\
x_2 + \alpha_{23}x_3 + \cdots + \alpha_{2r}x_r &=& -\alpha_{2,r+1}x_{r+1} - \cdots - \alpha_{2n}x_n + \beta_2 \\
\cdots\cdots\cdots\cdots && \cdots\cdots\cdots\cdots \\
x_r &=& -\alpha_{r,r+1}x_{r+1} - \cdots - \alpha_{rn}x_n + \beta_r
\end{array}$$

In this form it is clear that $x_{r+1}, x_{r+2}, \ldots, x_n$ can be given arbitrary values, say

$$x_{r+1} = \lambda_1,\ x_{r+2} = \lambda_2,\ \ldots,\ x_n = \lambda_{n-r}$$

Substituting these values into the last equation in the above system, we obtain x_r immediately. Then, substituting for $x_r, x_{r+1}, \ldots, x_n$ in the next to the last equation, we obtain x_{r-1}, and so on, step by step until each of the x's is determined. Finally, when the expressions for $x_1, x_2, \ldots, x_r$ are simplified by collecting terms on $\lambda_1, \lambda_2, \ldots, \lambda_{n-r}$ we obtain expressions of the form

$$\begin{array}{rcl}
x_1 &=& \lambda_1 c_{11} + \lambda_2 c_{12} + \cdots + \lambda_{n-r}c_{1,n-r} + \gamma_1 \\
\cdots && \cdots\cdots\cdots\cdots \\
x_r &=& \lambda_1 c_{r1} + \lambda_2 c_{r2} + \cdots + \lambda_{n-r}c_{r,n-r} + \gamma_r \\
x_{r+1} &=& \lambda_1 \\
x_{r+2} &=& \qquad\lambda_2 \\
\cdots && \cdots\cdots\cdots \\
x_n &=& \qquad\qquad\lambda_{n-r}
\end{array}$$

or

$$X = \left\| \begin{matrix} x_1 \\ \cdot \\ x_r \\ x_{r+1} \\ x_{r+2} \\ \cdot \\ x_m \end{matrix} \right\| = \lambda_1 \left\| \begin{matrix} c_{11} \\ \cdot \\ c_{r1} \\ 1 \\ 0 \\ \cdot \\ 0 \end{matrix} \right\| + \lambda_2 \left\| \begin{matrix} c_{12} \\ \cdot \\ c_{r2} \\ 0 \\ 1 \\ \cdot \\ 0 \end{matrix} \right\| + \cdots + \lambda_{n-r} \left\| \begin{matrix} c_{1,n-r} \\ \cdot \\ c_{r,n-r} \\ 0 \\ 0 \\ \cdot \\ 1 \end{matrix} \right\| + \left\| \begin{matrix} \gamma_1 \\ \cdot \\ \gamma_r \\ 0 \\ 0 \\ \cdot \\ 0 \end{matrix} \right\|$$

The $n - r$ vectors which are multiplied respectively by $\lambda_1, \lambda_2, \ldots, \lambda_{n-r}$ depend only on the a_{ij}'s in the original system and are, in fact, solution vectors of the related homogeneous system $AX = O$. The vector

$$\left\| \begin{matrix} \gamma_1 \\ \cdot \\ \gamma_r \\ 0 \\ 0 \\ \cdot \\ 0 \end{matrix} \right\|$$

depends not only on the a_{ij}'s but also on the b_i's and is clearly a particular solution of the nonhomogeneous system $AX = B$. By Theorem 8, the $n - r$ vectors which are multiplied by the λ's are linearly independent. Hence, the related homogeneous system $AX = O$ has $n - r$ linearly independent solutions, and the complete solution of the nonhomogeneous system $AX = B$ contains $n - r$ independent arbitrary constants, as asserted.

EXAMPLE 3

Find a complete solution of the system

$$\begin{aligned} x_1 + 2x_2 + x_3 - x_4 + 2x_5 &= 2 \\ x_1 + 4x_2 + 5x_3 - 3x_4 + 8x_5 &= -2 \\ -2x_1 - x_2 + 4x_3 - x_4 + 5x_5 &= -10 \\ 3x_1 + 7x_2 + 5x_3 - 4x_4 + 9x_5 &= 4 \end{aligned}$$

Applying the Gauss reduction, we obtain successively

$$\begin{aligned} x_1 + 2x_2 + x_3 - x_4 + 2x_5 &= 2 \\ 2x_2 + 4x_3 - 2x_4 + 6x_5 &= -4 \\ 3x_2 + 6x_3 - 3x_4 + 9x_5 &= -6 \\ x_2 + 2x_3 - x_4 + 3x_5 &= -2 \end{aligned} \qquad \text{and} \qquad \begin{aligned} x_1 + 2x_2 + x_3 - x_4 + 2x_5 &= 2 \\ x_2 + 2x_3 - x_4 + 3x_5 &= -2 \\ 0 &= 0 \\ 0 &= 0 \end{aligned}$$

Hence, we have the solutions

$$\begin{aligned} x_1 &= -2x_2 - x_3 + x_4 - 2x_5 + 2 \\ x_2 &= - 2x_3 + x_4 - 3x_5 - 2 \\ x_3 &= x_3 \\ x_4 &= x_4 \\ x_5 &= x_5 \end{aligned}$$

or, taking $x_3 = \lambda_1$, $x_4 = \lambda_2$, $x_5 = \lambda_3$,

$$\begin{aligned} x_1 &= 3\lambda_1 - \lambda_2 + 4\lambda_3 + 6 \\ x_2 &= -2\lambda_1 + \lambda_2 - 3\lambda_3 - 2 \\ x_3 &= \lambda_1 \\ x_4 &= \lambda_2 \\ x_5 &= \lambda_3 \end{aligned}$$

A complete solution of the original system is, therefore,

$$X = \lambda_1 \left\| \begin{matrix} 3 \\ -2 \\ 1 \\ 0 \\ 0 \end{matrix} \right\| + \lambda_2 \left\| \begin{matrix} -1 \\ 1 \\ 0 \\ 1 \\ 0 \end{matrix} \right\| + \lambda_3 \left\| \begin{matrix} 4 \\ -3 \\ 0 \\ 0 \\ 1 \end{matrix} \right\| + \left\| \begin{matrix} 6 \\ -2 \\ 0 \\ 0 \\ 0 \end{matrix} \right\|$$

where λ_1, λ_2, λ_3 are arbitrary scalars

The existence of a solution for the nonhomogeneous system implies that the coefficient matrix and the augmented matrix

$$A = \left\| \begin{matrix} 1 & 2 & 1 & -1 & 2 \\ 1 & 4 & 5 & -3 & 8 \\ -2 & -1 & 4 & -1 & 5 \\ 3 & 7 & 5 & -4 & 9 \end{matrix} \right\| \qquad \text{and} \qquad \|AB\| = \left\| \begin{matrix} 1 & 2 & 1 & -1 & 2 & 2 \\ 1 & 4 & 5 & -3 & 8 & -2 \\ -2 & -1 & 4 & -1 & 5 & -10 \\ 3 & 7 & 5 & -4 & 9 & 4 \end{matrix} \right\|$$

have the same rank. The fact that the complete solution contains three arbitrary constants implies that the common value of the rank of the two matrices is 2, since, according to the last theorem,

$$\begin{aligned}&(\text{Number of arbitrary constants in complete solution})\\ &\qquad = (\text{number of unknowns}) - (\text{common value of rank})\end{aligned}$$

It is, of course, not difficult to verify that A and $\|AB\|$ are both of rank 2. The vector

$$X_P = \begin{Vmatrix} 6 \\ -2 \\ 0 \\ 0 \\ 0 \end{Vmatrix}$$

is a particular solution of the given nonhomogeneous equation $AX = B$. The vectors

$$X_1 = \begin{Vmatrix} 3 \\ -2 \\ 1 \\ 0 \\ 0 \end{Vmatrix} \qquad X_2 = \begin{Vmatrix} -1 \\ 1 \\ 0 \\ 1 \\ 0 \end{Vmatrix} \qquad \text{and} \qquad X_3 = \begin{Vmatrix} 4 \\ -3 \\ 0 \\ 0 \\ 1 \end{Vmatrix}$$

are three linearly independent solutions of the related homogeneous equation $AX = O$. That $\lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 + X_P$ is a complete solution of the given system follows, of course, from Theorem 6.

As the last example illustrated, the Gauss reduction provides a practical method for solving systems of simultaneous linear equations in the general case. However, in several important special cases there are other methods which are sometimes more convenient. Specifically, we have the following pair of theorems:

THEOREM 7 (CRAMER'S RULE)
If the coefficient matrix A of a system $AX = B$ of n linear equations in n unknowns is nonsingular, the system has the unique solution

$$x_1 = \frac{|D_1|}{|A|}, \qquad x_2 = \frac{|D_2|}{|A|}, \qquad \cdots, \qquad x_n = \frac{|D_n|}{|A|}$$

where D_i is the matrix obtained from A by replacing the ith column of A by the column vector B.

PROOF Since the matrix of coefficients A is nonsingular by hypothesis, it follows that its inverse A^{-1} exists. Hence, premultiplying the given equation $AX = B$ by A^{-1}, we obtain

$$X = A^{-1}B$$

and direct substitution confirms that this actually is a solution. Now, in the column vector $A^{-1}B$, the element in the ith row is simply the scalar product of the ith row vector of A^{-1} and B itself, that is,

$$\frac{A_{1i}b_1 + A_{2i}b_2 + \cdots + A_{ni}b_n}{|A|}$$

Moreover, the numerator of this fraction is just the expansion, in terms of the ith column, of the determinant of the matrix D_i obtained from A by replacing the ith column of A by the column vector B. Hence

$$(4) \qquad x_i = \frac{|D_i|}{|A|} \qquad \text{as asserted}$$

Since A is nonsingular, the rank of the coefficient matrix and the rank of the augmented matrix are both equal to n Hence, according to Theorem 6, there can be no arbitrary constant in any complete solution, and the solution we have found is the only one.

If in the (n,n) system $AX = B$ the vector B is zero, that is, if $b_1 = b_2 = \cdots = b_n = 0$, then, clearly, each determinant $|D_i|$ contains a column consisting entirely of zeros and hence is zero. If $|A| \neq 0$, it therefore follows from (4) that $x_i = 0$ for all values of i, or, in other words, that only a **trivial solution** is possible On the other hand, if $B = O$, the coefficient matrix and the augmented matrix have the same rank, and if $|A| = 0$ the common value of these ranks is at most $r = n - 1$; that is, $r \leqq n - 1$ Hence $n - r$ is at least as much as 1, and, by Theorem 6, the equation $AX = O$ has at least one nontrivial solution vector Thus we have established the following important corollary of Theorem 7.

COROLLARY 1

A homogeneous system of n linear equations in n unknowns $AX = O$ has a nontrivial solution, i e , a solution other than $x_1 = x_2 = \quad = x_n = 0$, if and only if the determinant of the coefficients $|A|$ is equal to zero

More specifically, when the rank of the coefficient matrix of a homogeneous system of n linear equations in n unknowns is $n - 1$, we have the following useful result·

THEOREM 8

If the coefficient matrix of a homogeneous system of n linear equations in n unknowns $AX = O$ is of rank $n - 1$ and if the submatrix obtained from A by omitting the kth row is also of rank $n - 1$, then a complete solution of the given system is

$$x_i = cA_{ki} \qquad i = 1, 2, \quad , n$$

where c is an arbitrary constant.

PROOF Since the rank of the $(n - 1, n)$ matrix remaining when the kth row is deleted from A is $n - 1$, it follows that at least one of the cofactors A_{ki} of the kth row is different from zero. Hence, not all of the values $x_i = cA_{ki}$ are zero To verify that these values do indeed satisfy $AX = O$, we need only substitute them into the general equation of the system, namely,

$$\sum_{i=1}^{n} a_{ji}x_i = 0 \qquad j = 1, 2, \ldots, n$$

and verify that it is satisfied. Doing this, and using Corollary 1, Theorem 11, Sec. 10 1, to simplify the result, we have

$$\sum_{i=1}^{n} a_{ji}(cA_{ki}) = c\sum_{i=1}^{n} a_{ji}A_{ki} = \begin{cases} 0 & j \neq k \\ |A| & j = k \end{cases}$$

Thus, since $|A| = 0$, by hypothesis, it follows that each equation is satisfied by the given values Finally, since the rank of both the coefficient matrix and the augmented matrix is $r = n - 1$, it follows by Theorem 7 that the system has just one independent solution vector Hence the solution given by the formula of the theorem is a complete solution, as asserted

EXAMPLE 4

Find a complete solution of the system

$$\begin{aligned} x_1 - 2x_2 + x_3 + 3x_4 &= 0 \\ 2x_1 + 2x_2 - x_3 + x_4 &= 0 \\ -x_1 - x_2 + 3x_3 + 2x_4 &= 0 \\ x_1 - 8x_2 - x_3 + 3x_4 &= 0 \end{aligned}$$

It is easy to verify that the determinant of the coefficients of this system, $|A|$, is zero but that the determinant of the (3,3) submatrix in the upper left hand corner of A is different from zero Thus the coefficient matrix A and the submatrix remaining when the last row is deleted from A are both of rank 3 Hence, according to the last theorem, the values of x which satisfy the given system of equations are proportional to the cofactors of the last row in $|A|$ Thus we have

$$= c\begin{vmatrix} -2 & 1 & 3 \\ 2 & -1 & 1 \\ -1 & 3 & 2 \end{vmatrix} = 20c \qquad x_2 = -c\begin{vmatrix} 1 & 1 & 3 \\ 2 & -1 & 1 \\ -1 & 3 & 2 \end{vmatrix} = -5c$$

$$x_3 = c\begin{vmatrix} 1 & -2 & 3 \\ 2 & 2 & 1 \\ -1 & -1 & 2 \end{vmatrix} = 15c \qquad x_4 = -c\begin{vmatrix} 1 & -2 & 1 \\ 2 & 2 & -1 \\ -1 & -1 & 3 \end{vmatrix} = 15c$$

or, setting $5c = k$,

$$x_1 = 4k \qquad x_2 = -k \qquad x_3 = 3k \qquad x_4 = -3k$$

In view of Theorems 7 and 8, it is clearly desirable to have some convenient criterion for determining when a determinant is different from zero. One useful one is provided by the following theorem, whose proof is an interesting application of Corollary 1, Theorem 7

THEOREM 9

If in each row of a determinant the absolute value of the element on the principal diagonal is greater than the sum of the absolute values of the remaining elements in that row, the value of the determinant is different from zero

PROOF Let $|A|$ be an $n \times n$ determinant in which, for each value of i,

$$(5) \qquad |a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|†$$

---

† This property of the absolute values of the elements of a determinant is known as **diagonal dominance** and is quite important in the solution of systems of linear equations by iterative methods.

To prove that $|A|$ is not equal to zero, let us assume the contrary. Now, if $|A| = 0$, then by Corollary 1, Theorem 7, the equations

$$\begin{array}{ll} & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ (6) & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = 0 \end{array}$$

have a nontrivial solution. Let x_k be the component of maximum absolute value of such a solution vector. Then from the kth equation of the system (6) we have, by transposing,

$$a_{kk}x_k = -\sum_{\substack{j=1 \\ j\neq k}}^{n} a_{kj}x_j$$

Hence, taking absolute values,

$$|a_{kk}|\ |x_k| \leqq \sum_{\substack{j=1 \\ j\neq k}}^{n} |a_{kj}|\ |x_j| \leqq \sum_{\substack{j=1 \\ j\neq k}}^{n} |a_{kj}| \cdot |x_k|$$

and therefore, dividing out $|x_k|$,

$$|a_{kk}| \leqq \sum_{\substack{j=1 \\ j\neq k}}^{n} |a_{kj}|$$

But this contradicts some one of the inequalities (5), which hold, by hypothesis. Therefore we must abandon the assumption that $|A| = 0$, and the theorem is established.

If a set of vectors has the property that it contains at least one subset of r vectors which are linearly independent, but all $r + 1$ vectors are linearly dependent, the set is said to be of **dimension** r. Using Theorem 6, it is not difficult to prove the following theorem about the dimension of a set of vectors:

THEOREM 10

A set of row (column) vectors $X_1, X_2, \ldots, X_n$ is of dimension r if and only if the matrix

$$\left\|\begin{matrix} X_1 \\ X_2 \\ \cdot \\ X_n \end{matrix}\right\| \qquad (\|X_1X_2 \cdots X_n\|)$$

is of rank r.

COROLLARY 1

If $X_1, X_2, \ldots, X_n$ are n linearly independent vectors each having n components, then any vector B with n components can be expressed as a linear combination of the X's.

In Sec. 10.4 we observed that what we called simply the *rank* of a matrix is sometimes referred to as the *determinant rank*. This permits one to distinguish between the rank, as we defined it, and the **row rank** and **column rank,** which are, by definition, the maximum number of linearly independent row and column vectors, respectively, in the matrix. However, the necessity for distinguishing between the three definitions of rank is eliminated by the following theorem, whose proof follows immediately from Theorem 10:

THEOREM 11

For any matrix A, the determinant rank, the row rank, and the column rank are all equal.

Another interesting consequence of Theorem 10 is contained in the following theorem:

THEOREM 12

If A and B are conformable matrices of rank r and ρ, respectively, the rank of the product AB is equal to or less than the smaller of the numbers r and ρ.

PROOF Let A be an (m,p) matrix, let B be a (p,n) matrix, and let the rows of A be $A_1, A_2, \ldots, A_m$. Then the ith row vector in the product AB (see Exercise 7, Sec. 10.2) is

$$(7) \qquad A_iB$$

Now, by hypothesis, the rank of A is r. Hence, by Theorem 10, A contains exactly r linearly independent row vectors, which, without loss of generality, we can take to be the first r, namely, $A_1, A_2, \ldots, A_r$. Hence, for $i = r+1, r+2, \ldots, m$, the row A_i must be expressible as a linear combination of the first r rows:

$$A_i = \lambda_{1i}A_1 + \lambda_{2i}A_2 + \cdots + \lambda_{ri}A_r \qquad i = r+1, r+2, \ldots, m$$

Therefore, substituting into (7), we find that, for $i = r+1, r+2, \ldots, m$, the ith row vector of the product AB is

$$(\lambda_{1i}A_1 + \lambda_{2i}A_2 + \cdots + \lambda_{ri}A_r)B = \lambda_{1i}A_1B + \lambda_{2i}A_2B + \cdots + \lambda_{ri}A_rB$$

But this shows that each row of AB after the rth is a linear combination of the first r rows, which, in turn, proves that AB contains at most r linearly independent row vectors and, hence, is of rank at most r. A similar argument, using a column partition of B, shows that the rank of AB is at most equal to ρ. Therefore, the rank of AB is at most equal to the smaller of the numbers r and ρ, as asserted. If A and B are also conformable in the order BA, it is clear that the rank of BA is also equal to or less than the smaller of the pair (r,ρ).

The estimate for the rank of the product AB provided by the last theorem can be supplemented with the following result.*

THEOREM 13

If A is an (m,p) matrix of rank r and if B is a (p,n) matrix of rank ρ, the rank of the product AB is equal to or greater than $r + \rho - p$.

---

* Both this result and Theorem 12 are due to the English mathematician J. J. Sylvester (1814–1897) and are known together as *Sylvester's law of nullity*. A proof of Theorem 13 can be found in L. Mirsky, "Linear Algebra," p. 162, Oxford Book Company, Inc., New York, 1955.

As we shall see in later sections, a set of vectors is usually much more convenient to work with if the vectors, in addition to being linearly independent, are also **orthonormal**, that is, are of unit length and mutually orthogonal. A general set of r linearly independent vectors will ordinarily not possess the property of orthonormality, but, by an important procedure known as the **Schmidt orthogonalization process,** it is always possible to determine linear combinations of r linearly independent vectors which will be orthonormal as well as independent. Let $V_1, V_2, \ldots, V_n$ be n linearly independent vectors, and let us choose any one of them, say V_1, and reduce it to a unit vector by dividing it by its length $\sqrt{V_1^TV_1}$. This gives us the first vector of our orthonormal set

$$U_1 = \frac{V_1}{\sqrt{V_1^TV_1}}$$

We now choose any member of the original set except V_1, say V_2, and write

$$W_2 = V_2 - c_1U_1$$

where c_1 is a constant to be determined so that W_2 is orthogonal to U_1. This, of course, requires that

$$U_1^TW_2 = U_1^T(V_2 - c_1U_1) = 0$$

From this, since $U_1^TU_1 = 1$, we have

$$c_1 = U_1^TV_2 \qquad \text{and} \qquad W_2 = V_2 - (U_1^TV_2)U_1$$

We now convert W_2 to a unit vector by dividing it by its length, getting

$$U_2 = \frac{W_2}{\sqrt{W_2^TW_2}} = \frac{V_2 - (U_1^TV_2)U_1}{\sqrt{[V_2 - (U_1^TV_2)U_1]^T[V_2 - (U_1^TV_2)U_1]}}$$

Next we choose any member of the original set except V_1 and V_2, say V_3, and write

$$W_3 = V_3 - d_1U_1 - d_2U_2$$

where d_1 and d_2 are constants to be determined so that W_3 will be orthogonal to both U_1 and U_2. This gives us the two conditions

$$U_1^TW_3 = U_1^T(V_3 - d_1U_1 - d_2U_2) = U_1^TV_3 - d_1 = 0$$
$$U_2^TW_3 = U_2^T(V_3 - d_1U_1 - d_2U_2) = U_2^TV_3 - d_2 = 0$$

Hence,

$$d_1 = U_1^TV_3 \qquad \text{and} \qquad d_2 = U_2^TV_3$$

and, therefore,

$$W_3 = V_3 - (U_1^TV_3)U_1 - (U_2^TV_3)U_2$$

W_3 is now normalized, giving us our third unit vector U_3; and the process is continued until the required set of orthonormal vectors is obtained from the original set. It is clear that the process can fail if and only if at some stage

$$W_k = V_k - \sum_{i=1}^{k-1} (U_i^T V_k) U_i = O$$

However, if $W_k = O$, this implies that

$$V_k = \sum_{i=1}^{k-1} (U_i^T V_k) U_i$$

which, replacing the U's by their expressions in terms of V_1, V_2, . . . , V_{k-1}, implies that V_k is either zero or else a linear combination of the preceding V's. Each of these contradicts the hypothesis that the V's are linearly independent, and hence cannot happen. An almost identical argument (or the result of Exercise 14) shows that the U's derived by the Schmidt process are linearly independent.

EXERCISES

1 Verify that $\sin x$ and $\cos x$ are linearly independent.

2 Are $\cos^2 x$, $\sin^2 x$, and $\cos 2x$ linearly independent? Why?

3 Show that, if 0 is included in a set of quantities, the members of the set are always linearly dependent.

4 Show that, if the quantities $Q_1, Q_2, \ldots, Q_n$ are linearly independent, the members of every subset of the Q's are also linearly independent. Is the converse true?

5 If A is a square matrix and if the equation $AX = O$ has k linearly independent solution vectors, show that the same is true of the system $A^TX = O$. Is this result true if A is not a square matrix?

6 Show that five or more 2×2 matrices are always linearly dependent. How many 3×3 matrices must we have before we can be sure they are linearly dependent? Why?

7 Verify that the vectors

$$X_1 = \left\Vert\begin{matrix}1\\1\\0\end{matrix}\right\Vert \qquad X_2 = \left\Vert\begin{matrix}1\\-1\\1\end{matrix}\right\Vert \qquad X_3 = \left\Vert\begin{matrix}2\\1\\3\end{matrix}\right\Vert \qquad X_4 = \left\Vert\begin{matrix}-1\\4\\-5\end{matrix}\right\Vert$$

are linearly dependent, and express each of the vectors as a linear combination of the other three.

8 What conditions must a, b, c, and d satisfy in order that the matrices $\left\Vert\begin{matrix}1 & 2\\-1 & 0\end{matrix}\right\Vert$, $\left\Vert\begin{matrix}2 & 3\\-2 & 1\end{matrix}\right\Vert$, and $\left\Vert\begin{matrix}a & b\\c & d\end{matrix}\right\Vert$ be linearly dependent?

9 Using the Gauss reduction, find a complete solution of each of the following systems.

a
$$\begin{aligned}x_1 + 2x_2 + 4x_3 - x_4 + 2x_5 &= 3\\ 3x_1 + 4x_2 + 5x_3 - x_4 - 2x_5 &= 7\\ x_1 + 3x_2 + 4x_3 + 5x_4 - x_5 &= 4\end{aligned}$$

b
$$\begin{aligned}x_1 + x_2 + x_3 - x_4 - x_5 &= 2\\ x_1 + 2x_2 + 4x_3 - x_4 + 5x_5 &= 3\\ 3x_1 + 4x_2 + 5x_3 - x_4 - 2x_5 &= 7\\ x_1 + 3x_2 + 4x_3 + 5x_4 - x_5 &= 4\\ 2x_1 + 5x_2 + 8x_3 + 4x_4 + x_5 &= 7\\ x_1 - x_2 - 2x_3 - 7x_4 - x_5 &= 0\end{aligned}$$

10 Using Cramer's rule, solve each of the following systems:

a
$$\begin{array}{rcrcrcr} x_1 & - & x_2 & + & 2x_3 & = & -5 \\ -x_1 & & & + & 3x_3 & = & 0 \\ 2x_1 & + & x_2 & & & = & 1 \end{array}$$

b
$$\begin{array}{rcrcrcrcr} x_1 & - & x_2 & + & 2x_3 & + & x_4 & = & -5 \\ -x_1 & & & + & 3x_3 & + & 2x_4 & = & 0 \\ 2x_1 & + & x_2 & & & - & x_4 & = & 1 \\ 2x_1 & + & 2x_2 & + & x_3 & + & 3x_4 & = & -1 \end{array}$$

11 Using Theorem 8, solve each of the following systems:

a
$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & 3x_3 & = & 0 \\ 2x_1 & + & 3x_2 & - & x_3 & = & 0 \\ 4x_1 & - & x_2 & + & 5x_3 & = & 0 \end{array}$$

b
$$\begin{array}{rcrcrcrcr} x_1 & - & 2x_2 & + & x_3 & - & 3x_4 & = & 0 \\ 2x_1 & + & x_2 & - & 3x_3 & + & x_4 & = & 0 \\ 3x_1 & + & 3x_2 & - & 2x_3 & + & x_4 & = & 0 \end{array}$$

12 Determine the values of λ, if any, for which each of the following systems has a nontrivial solution, and find such solutions when they exist.

a
$$\begin{array}{rcrcrcr} \lambda x_1 & - & 2x_2 & + & x_3 & = & 0 \\ \lambda x_1 & + & (1-\lambda)x_2 & + & x_3 & = & 0 \\ 2x_1 & - & x_2 & + & 2\lambda x_3 & = & 0 \end{array}$$

b
$$\begin{array}{rcrcrcr} (5-\lambda)x_1 & + & 4x_2 & - & 2x_3 & = & 0 \\ 4x_1 & + & (5-\lambda)x_2 & - & 2x_3 & = & 0 \\ -2x_1 & - & 2x_2 & + & (3-2\lambda)x_3 & = & 0 \end{array}$$

13 If the rows of a matrix are linearly dependent, are the columns necessarily linearly dependent?

14 Show that, if the vectors of a set are mutually orthogonal, they are linearly independent.

15 Show that, if r is the maximum number of linearly independent quantities in the set $Q_1, Q_2, \dots, Q_n$ and if $Q_1, Q_2, \dots, Q_r$ are linearly independent, then $Q_{r+1}, Q_{r+2}, \dots, Q_n$ can each be expressed as a linear combination of $Q_1, Q_2, \dots, Q_r$.

16 If the quantities $Q_1, Q_2, \dots, Q_n$ are such that $Q_{r+1}, Q_{r+2}, \dots, Q_n$ can each be expressed as a linear combination of $Q_1, Q_2, \dots, Q_r$, show that at most r of the Q's can be linearly independent.

17 Show that any (3,4) matrix of rank 2 can be written as the product of a (3,2) matrix of rank 2 and a (2,4) matrix of rank 2. Of what general theorem do you think this is a special case?

18 Prove Theorem 10.

19 Prove Corollary 1, Theorem 10.

20 Using the Schmidt process, construct a set of orthonormal vectors from the vectors in each of the following sets:

a $V_1 = \|1,2,2\|$ $V_2 = \|1,4,0\|$ $V_3 = \|2,0,1\|$

b $V_1 = \|1,1,0\|$ $V_2 = \|1,0,1\|$ $V_3 = \|0,1,1\|$

c $V_1 = \|1,1,1,1\|$ $V_2 = \|0,1,2,2\|$ $V_3 = \|0,0,1,1\|$

21 Prove that, if an $(n,n+1)$ matrix A contains a column of elements which are not all zero and if every nth-order determinant in A which contains this column vanishes, then the rank of A is less than n. (Hint: Expand each of the vanishing determinants in terms of the elements in their common column, consider the determinant of the resulting system of equations, and use the result of Exercise 8, Sec. 10.3.)

22 Prove that n vectors $V_1, V_2, \dots, V_n$ are linearly dependent if and only if the so-called **Gram determinant**, or **Gramian**,

$$\begin{vmatrix} V_1^TV_1 & V_1^TV_2 & \cdots & V_1^TV_n \\ V_2^TV_1 & V_2^TV_2 & \cdots & V_2^TV_n \\ \cdots & \cdots & \cdots & \cdots \\ V_n^TV_1 & V_n^TV_2 & \cdots & V_n^TV_n \end{vmatrix} \text{ is equal to zero}$$

23 If A is a square matrix, p a positive integer, and X a vector such that $A^pX \neq O$ but $A^{p+1}X = O$, show that the vectors $X, AX, A^2X, \dots, A^pX$ are linearly independent.

24 **a** Let A and B be matrices conformable in the order AB. Prove that the rank of AB is equal to the rank of B if and only if $BX = O$ for every vector X such that $ABX = O$.

b If A is a square matrix and if p is a positive integer such that A^p and A^{p+1} have the same rank r, show that A^{p+2}, A^{p+3}, A^{p+4}, . . . also are of rank r

25 Prove that, if A is an (n,n) matrix, then A^n, A^{n+1}, A^{n+2}, . . . all have the same rank

10.6

Matric differential equations

In Sec. 3.3 we saw that the ideas of *complementary function* and *particular integral* which we developed for single linear differential equations could easily be extended to systems of linear differential equations. This analogy is especially striking when we regard a system of linear differential equations with constant coefficients as a single matric equation, much as we regarded a system of linear algebraic equations as a single matric equation in the last section Moreover, the procedure for handling systems of equations when the characteristic equation has repeated or complex roots or when a term on the right-hand side duplicates a term in the complementary function is best described in the language of matrices. Hence we shall conclude this chapter with a brief discussion of matric differential equations.

Let the system we are given be

$$\begin{aligned} p_{11}(D)x_1 + p_{12}(D)x_2 + \cdots + p_{1n}(D)x_n &= f_1(t) \\ p_{21}(D)x_1 + p_{22}(D)x_2 + \cdots + p_{2n}(D)x_n &= f_2(t) \\ \cdots\cdots\cdots\cdots\cdots\cdots & \cdots \\ p_{n1}(D)x_1 + p_{n2}(D)x_2 + \cdots + p_{nn}(D)x_n &= f_n(t) \end{aligned} \tag{1}$$

where the p_{ij}'s are polynomials in the operator D with constant coefficients If we define the matrices

$$P(D) = \left\| \begin{matrix} p_{11}(D) & p_{12}(D) & \cdots & p_{1n}(D) \\ p_{21}(D) & p_{22}(D) & \cdots & p_{2n}(D) \\ \cdot & \cdot & \cdot & \cdot \\ p_{n1}(D) & p_{n2}(D) & \cdots & p_{nn}(D) \end{matrix} \right\| \qquad X = \left\| \begin{matrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{matrix} \right\| \qquad F(t) = \left\| \begin{matrix} f_1(t) \\ f_2(t) \\ \cdot \\ f_n(t) \end{matrix} \right\|$$

the system (1) can be written in the compact form

$$P(D)X = F(t) \tag{2}$$

The associated homogeneous equation is, of course,

$$P(D)X = O \tag{3}$$

The first step in finding the complementary function of Eq. (2) is to assume that solutions of Eq (3) exist in the form

$$X = Ae^{mt}$$

where the scalar m and the column matrix of constants A have yet to be determined. [The expressions

$$x = ae^{mt} \qquad y = be^{mt} \qquad z = ce^{mt}$$

in Eq. (3), Sec. 3.3, are, of course, just the scalar form of this assumption for the special case $n = 3$.] Since

$$D^r(e^{mt}) = m^r e^{mt}$$

it follows that, if we substitute the vector $X = Ae^{mt}$ into the homogeneous equation (3), we obtain just

$$P(m)Ae^{mt} = O$$

or, dividing out the nonvanishing scalar factor e^{mt},

$$P(m)A = O \tag{4}$$

This is the matric equivalent of the algebraic system in Eq. (4), Sec. 3.3, which we obtained in our scalar treatment of the specific system of differential equations we considered in that section. Now by Corollary 1, Theorem 7, Sec. 10.5, Eq. (4) will have a nontrivial solution if and only if

$$|P(m)| = 0 \tag{5}$$

and for each root m_j of this equation there will be a solution vector A_j of (4) determined to within an arbitrary scalar factor k_j. If the characteristic equation (5) is of degree N and if its roots $\{m_j\}$ are all distinct, a complete solution of Eq. (3)—and the complementary function of Eq. (2)—is then

$$X = k_1A_1e^{m_1t} + k_2A_2e^{m_2t} + \cdots + k_Ne^{m_Nt}$$

This we recognize as the matric equivalent of the scalar system (9), Sec. 3.3, with $N = 3$ and

$$A_1 = \left\|\begin{matrix} -1 \\ 2 \\ -1 \end{matrix}\right\| \qquad A_2 = \left\|\begin{matrix} 3 \\ -1 \\ -1 \end{matrix}\right\| \qquad A_3 = \left\|\begin{matrix} 9 \\ -6 \\ -1 \end{matrix}\right\|$$

As in the case of a single scalar differential equation, if the set of roots $\{m_j\}$ includes one or more pairs of conjugate complex roots, it is desirable to reduce the corresponding complex exponential solution to a purely real form. To see how this can be accomplished, let $p \pm iq$ be a pair of conjugate complex roots of Eq. (5), and let A be a particular solution vector of (4) corresponding to the root $m = p + iq$, that is, let

$$P(m)A \equiv P(p + iq)A = O$$

Then, since all the coefficients in (4) are real, it follows by taking conjugates throughout the system that

$$P(\bar{m})\bar{A} \equiv P(p - iq)\bar{A} = O$$

Thus $\bar{A}$ is a solution vector corresponding to the conjugate root

$$\bar{m} = p - iq$$

and, therefore, we have the two particular solutions of Eq. (3),

$$Ae^{(p+iq)t} \qquad \text{and} \qquad \bar{A}e^{(p-iq)t}$$

By combining these as follows and applying the Euler formulas, we obtain two independent real solutions

$$\frac{Ae^{(p+iq)t} + \bar{A}e^{(p-iq)t}}{2} = e^{pt}\left(\frac{A+\bar{A}}{2}\cos qt - \frac{A-\bar{A}}{2i}\sin qt\right)$$

(6a) $$= e^{pt}[\mathcal{R}(A)\cos qt - \mathcal{I}(A)\sin qt]$$

$$\frac{Ae^{(p+iq)t} - \bar{A}e^{(p-iq)t}}{2i} = e^{pt}\left(\frac{A-\bar{A}}{2i}\cos qt + \frac{A+\bar{A}}{2}\sin qt\right)$$

(6b) $$= e^{pt}[\mathcal{I}(A)\cos qt + \mathcal{R}(A)\sin qt]$$

where $\mathcal{R}(A)$ and $\mathcal{I}(A)$ denote the column matrices whose components are, respectively, the real parts of the components of A and the imaginary parts of the components of A. In many cases this method of determining the necessary relations among the coefficients of solutions of (3) of the form

$$x_j = e^{pt}(a_j \cos qt + b_j \sin qt)$$

is simpler than the alternative process of substituting these expressions into the original differential equations, collecting terms, and equating the resulting coefficients to zero.

If $|P(m)| = 0$ has a double root, say $m = r$, we proceed very much as in the case of a single differential equation. If A is a solution of the equation $P(r)A = O$, then, of course,

$$Ae^{rt}$$

is one solution of (3). However, as a second independent solution we must try not Bte^{rt}, as strict analogy with the scalar case would suggest, but rather

(7) $$B_1te^{rt} + B_2e^{rt}$$

The term B_2e^{rt} must be retained in the matric case because in general the matrix B_2 will not be a scalar multiple of A, and hence, in constructing the complete solution, the term B_2e^{rt} cannot be absorbed in the term Ae^{rt}, as is necessarily the case for a single scalar differential equation. It can be shown, however, that to within an arbitrary scalar factor the matrix B_1 is the same as A. Hence, after (7) has been substituted into the homogeneous system (3), it is only necessary to solve for the ratios of the components of the matrix B_2. Similar observations hold for roots of (5) of higher multiplicity. Thus, for a k-fold root r, the appropriate solutions are not Ae^{rt}, Bte^{rt}, , $Kt^{k-1}e^{rt}$ but rather

$$Ae^{rt}$$

$$B_1te^{rt} + B_2e^{rt}$$

$$\cdot \quad \cdot \quad \cdot$$

$$K_1t^{k-1}e^{rt} + K_2t^{k-2}e^{rt} + \qquad + K_ke^{rt}$$

In this case, to within arbitrary scalar factors, the matrices A, B_1, , K_1 are identical.

EXAMPLE 1

Find a complete solution of the system

$$(D^2 + D + 8)x_1 + (D^2 + 6D + 3)x_2 = 0$$
$$(D + 1)x_1 + (D^2 + 1)x_2 = 0$$

In this case the characteristic equation (5) is

$$\begin{vmatrix} (m^2 + m + 8) & (m^2 + 6m + 3) \\ (m + 1) & (m^2 + 1) \end{vmatrix} = m^4 + 2m^2 - 8m + 5 = 0$$

with roots 1, 1, $-1 \pm 2i$ For the root $-1 + 2i$, Eq (4) becomes

$$\begin{Vmatrix} (-1 + 2i)^2 + (-1 + 2i) + 8 & (-1 + 2i)^2 + 6(-1 + 2i) + 3 \\ (-1 + 2i) + 1 & (-1 + 2i)^2 + 1 \end{Vmatrix} \begin{Vmatrix} a_1 \\ a_2 \end{Vmatrix} = \begin{Vmatrix} 0 \\ 0 \end{Vmatrix}$$

or

$$\begin{Vmatrix} (4 - 2i) & (-6 + 8i) \\ 2i & (-2 - 4i) \end{Vmatrix} \begin{Vmatrix} a_1 \\ a_2 \end{Vmatrix} = \begin{Vmatrix} 0 \\ 0 \end{Vmatrix}$$

This is equivalent to the two scalar equations

$$(4 - 2i)a_1 + (-6 + 8i)a_2 = 0$$
$$2ia_1 - (2 + 4i)a_2 = 0$$

Since $m = -1 + 2i$ is a root of the characteristic equation (5), these two equations are dependent, and the ratio of a_1 to a_2 can be found equally well from either of them Using the second, since it is a little simpler, we therefore have

$$\frac{a_1}{a_2} = \frac{1 + 2i}{i} \qquad \text{or} \qquad A \equiv \begin{Vmatrix} a_1 \\ a_2 \end{Vmatrix} = \begin{Vmatrix} 1 + 2i \\ i \end{Vmatrix}$$

Hence,

$$\mathcal{R}(A) = \begin{Vmatrix} 1 \\ 0 \end{Vmatrix} \qquad \text{and} \qquad \mathcal{I}(A) = \begin{Vmatrix} 2 \\ 1 \end{Vmatrix}$$

and thus from (6) we have the two particular solutions

$$X_1 = e^{-t}\left(\begin{Vmatrix} 1 \\ 0 \end{Vmatrix} \cos 2t - \begin{Vmatrix} 2 \\ 1 \end{Vmatrix} \sin 2t\right) \qquad X_2 = e^{-t}\left(\begin{Vmatrix} 2 \\ 1 \end{Vmatrix} \cos 2t + \begin{Vmatrix} 1 \\ 0 \end{Vmatrix} \sin 2t\right)$$

For the repeated root $m = 1$, we have one solution of the form Be^t, where, from (4),

$$\begin{Vmatrix} 10 & 10 \\ 2 & 2 \end{Vmatrix} \begin{Vmatrix} b_1 \\ b_2 \end{Vmatrix} = \begin{Vmatrix} 0 \\ 0 \end{Vmatrix} \qquad \text{so we can take} \qquad B \equiv \begin{Vmatrix} b_1 \\ b_2 \end{Vmatrix} = \begin{Vmatrix} 1 \\ -1 \end{Vmatrix}$$

As a second solution we have, from (7),

$$C_1te^t + C_2e^t$$

or, since $C_1 = B$ (as we observed above, without proof),

$$\begin{Vmatrix} 1 \\ -1 \end{Vmatrix} te^t + \begin{Vmatrix} c_{12} \\ c_{22} \end{Vmatrix} e^t$$

Substituting this into the original system, we obtain two equations, each of which reduces to

$$2c_{12} + 2c_{22} = 1$$

Hence we can take*

$$c_{12} = 0 \qquad \text{and} \qquad c_{22} = \tfrac{1}{2}$$

The solutions associated with the double root $m = 1$ are, therefore,

$$X_3 = \begin{Vmatrix} 1 \\ -1 \end{Vmatrix} e^t \qquad \text{and} \qquad X_4 = \begin{Vmatrix} 1 \\ -1 \end{Vmatrix} te^t + \begin{Vmatrix} 0 \\ \tfrac{1}{2} \end{Vmatrix} e^t$$

---

* The most general choice, $c_{12} = \lambda$, $c_{22} = (1 - 2\lambda)/2$, leads to the same expression for X_4 plus a matrix proportional to X_3 which can be combined with X_3 when the complete solution is constructed.

The complete solution of the original system is now

$$X = \left\| \begin{matrix} x_1 \\ x_2 \end{matrix} \right\| = k_1X_1 + k_2X_2 + k_3X_3 + k_4X_4$$

or, in scalar form,

$$x_1 = e^{-t}[(k_1 + 2k_2)\cos 2t - (2k_1 - k_2)\sin 2t] + k_3e^t + k_4te^t$$

$$x_2 = e^{-t}[k_2\cos 2t - k_1\sin 2t] - (k_3 - \tfrac{1}{2}k_4)e^t - k_4te^t$$

To find a particular integral of the nonhomogeneous system (2), we proceed very much as in the case of a single scalar equation In fact, for vectors $F(t)$ which have only a finite number of independent derivatives and which do not duplicate vectors already in the complementary function, the results of Table 2.2, Sec 2 3, can be used without change, provided only that the arbitrary scalar constants appearing in the entries in the table be replaced by arbitrary constant vectors The trial solutions are then substituted into the nonhomogeneous system, and the arbitrary components of the coefficient vectors are determined to make the resulting equations identically true. The only significant difference between the scalar case and the matric case is that in the latter, when duplication occurs between a vector on the right of (2) and a vector in the complementary function, not only must the usual choice for a particular integral be multiplied by the lowest positive integral power of the independent variable which will eliminate the duplication but the products of the normal choice and all lower nonnegative integral powers of the independent variable must also be included in the actual choice

EXERCISES

Find a complete solution of each of the following systems·

1 $(D + 5)x + (D + 7)y = 2e^t$
$(2D + 1)x + (3D + 1)y = e^t$

2 $(D + 2)x + (D + 3)y = 2t + 4$
$(2D - 6)x + (3D - 4)y = -6t - 2$

3 $(D + 1)x + (D + 2)y = -e^t$
$(3D + 1)x + (4D + 7)y = -7e^t$

4 $(D + 1)x + (D + 2)y = -t + 1$
$(5D + 1)x + (6D + 3)y = -2t + 1$

5 $(2D + 1)x + (D + 2)y = 1$
$(D + 2)x + (D + 4)y = 2$

6 $(D + 1)x + (4D - 2)y = e^{-t}$
$(D + 2)x + (5D - 2)y = e^{-t}$

7 $(2D + 1)x + (D + 2)y = e^{-t}$
$(3D - 7)x + (3D + 1)y = 0$

8 $(2D + 1)x + (D + 2)y = \sin t$
$(3D + 1)x + (3D + 5)y = \cos t$

9 Show that $D^r(te^{mt}) = m^rte^{mt} + rm^{r-1}e^{mt}$ Hence show that

$$p(D)te^{mt} = p(m)te^{mt} + p'(m)e^{mt}$$

and $$P(D)te^{mt} = P(m)te^{mt} + P'(m)e^{mt}$$

where $p(D)$ is a polynomial in the operator D and $P(D)$ is a matrix whose elements are polynomials in D

10 Using the results of Exercise 9, show that, if m_1 is a double root of the characteristic equation $|P(m)| = 0$ and if $X_1 = Ae^{m_1t}$ is one solution of the system $P(D)X = O$, the coefficients in the second independent solution $X_2 = B_1te^{m_1t} + B_2e^{m_1t}$ satisfy the equations $P(m_1)B_1 = O$ and $P(m_1)B_2 + P'(m_1)B_1 = O$.

CHAPTER ELEVEN

Further Properties of Matrices

11.1

Quadratic forms

In this section we shall continue our study of matrices by introducing the important mathematical objects known as *quadratic forms, hermitian forms,* and *bilinear forms*.

By a **quadratic form** we mean a homogeneous second-degree expression in n variables of the form

$$\begin{array}{rlll} Q(x) = & a_{11}x_1^2 + 2a_{12}x_1x_2 + & \cdots & + 2a_{1n}x_1x_n \\ & \phantom{a_{11}x_1^2} + 2a_{22}x_2^2 \ + & \cdots & + 2a_{2n}x_2x_n \\ & \phantom{a_{11}x_1^2 + 2a_{22}x_2^2} \ + & \cdots & + \cdots \\ & & & + a_{nn}x_n^2 \end{array}$$

Usually the cross products are separated into two equal terms, and the whole expression is written in the more symmetric form

(1) $$\begin{array}{rllll} Q(x) = & a_{11}x_1^2 & + a_{12}x_1x_2 + & \cdots & + a_{1n}x_1x_n \\ & + a_{21}x_2x_1 & + a_{22}x_2^2 \ + & \cdots & + a_{2n}x_2x_n \\ & + \cdots & \cdots & \cdots & \cdots \\ & + a_{n1}x_nx_1 & + a_{n2}x_nx_2 + & \cdots & + a_{nn}x_n^2 \end{array}$$

where now, of course, $a_{ij} = a_{ji}$. If a quadratic form with real coefficients has the property that it is equal to or greater than (equal to or less than) zero for all real values of its variables, it is said to be **positive** (**negative**). A positive (negative) form which is zero only for the values $x_1 = x_2 = \cdots = x_n = 0$ is said to be **positive-definite** (**negative-definite**). Positive-definite and negative-definite forms are sometimes referred to collectively simply as **definite** forms. A positive (negative) form which is zero for real values other than $x_1 = x_2 = \cdots = x_n = 0$ is said to be **positive-semidefinite** (**negative-semidefinite**). A real quadratic form which can take on both positive and negative values is said to be **indefinite.** Examples of quadratic forms of each type are shown in the following table:

table 11.1

Type of quadratic form	Example
Positive-definite	$x_1^2 + x_2^2$
Negative-definite	$-(x_1^2 + x_2^2)$
Positive-semidefinite	$(x_1 - x_2)^2$
Negative-semidefinite	$-(x_1 - x_2)^2$
Indefinite	$x_1^2 - x_2^2$

If we define the matrices

$$X = \left\| \begin{matrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{matrix} \right\| \qquad \text{and} \qquad A = \left\| \begin{matrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{matrix} \right\|, \qquad a_{ij} = a_{ji}$$

it is clear from the definition of matric multiplication that the quadratic form (1) can be written in the compact form

$$Q(x) = X^T A X \qquad A \text{ symmetric} \tag{2}$$

In this notation A is called the **matrix of the quadratic form and** is said to be **positive-** or **negative-definite, semidefinite,** or **indefinite** according to the nature of $Q(x)$. $Q(x)$, in turn, is said to be **singular** or **nonsingular** according as A is singular or nonsingular, that is, according as $|A|$ is equal to zero or different from zero

If a quadratic form is definite, it is necessarily nonsingular, for we can write

$$\begin{aligned} Q(x) = {} & (a_{11}x_1 + \cdots + a_{1n}x_n)x_1 \\ & + (a_{21}x_1 + \cdots + a_{2n}x_n)x_2 \\ & + \cdots \cdots \cdots \cdots \cdots \\ & + (a_{n1}x_1 + \cdots + a_{nn}x_n)x_n \end{aligned}$$

and, if we suppose that $|A| = 0$, then the system of equations obtained by equating to zero the expressions in parentheses has a nontrivial solution (Corollary 1, Theorem 7, Sec 10 5), and for these values $Q(x)$ is obviously equal to zero, contrary to the hypothesis that it is definite The converse of this observation is not true, however; that is, a nonsingular quadratic form is not necessarily definite For instance, the form

$$x_1^2 - 2x_1x_2 + 2x_2^2 - x_3^2$$

is nonsingular, since the determinant of its matrix,

$$\begin{vmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & -1 \end{vmatrix} = -1$$

is different from zero; yet it is not definite, since it is equal to zero for the nontrivial values $x_1 = 1$, $x_2 = 0$, $x_3 = 1$ The complete criterion for the definiteness of a quadratic form is contained in the following theorem, for whose proof we must refer to texts on higher algebra *

HEOREM 1

A necessary and sufficient condition that the real quadratic form X^TAX be positive-definite (negative-definite) is that the quantities

$$a_{11}, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \quad \cdots, \quad \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \cdot & \cdots & \cdot \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

all be positive (alternate in sign, with a_{11} negative)

Clearly, equivalent sets of necessary and sufficient conditions can be obtained by first permuting the variables and then applying Theorem 1 This gives us the following somewhat more general theorems·

THEOREM 2

A necessary and sufficient condition that the real quadratic form X^TAX be positive-definite is that every principal minor of A be positive

THEOREM 3

A necessary and sufficient condition that the real quadratic form X^TAX be negative-definite is that every principal minor of A of odd order be negative and every principal minor of A of even order be positive

EXAMPLE 1

The quadratic form

$$\begin{Vmatrix} x_1 & x_2 & x_3 \end{Vmatrix} \begin{Vmatrix} 1 & 2 & -2 \\ 2 & 5 & -4 \\ -2 & -4 & 5 \end{Vmatrix} \begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix} = \begin{matrix} x_1^2 + 2x_1x_2 - 2x_1x_3 \\ +2x_2x_1 + 5x_2^2 - 4x_2x_3 \\ -2x_3x_1 - 4x_3x_2 + 5x_3^2 \end{matrix}$$

is positive-definite, since the three quantities

$$1 \quad \text{and} \quad \begin{vmatrix} 1 & 2 \\ 2 & 5 \end{vmatrix} = 1 \quad \text{and} \quad \begin{vmatrix} 1 & 2 & -2 \\ 2 & 5 & -4 \\ -2 & -4 & 5 \end{vmatrix} = 1$$

are all positive In fact, the quadratic form can be written equivalently as

$$(x_1 + 2x_2 - 2x_3)^2 + x_2^2 + x_3^2$$

---

* See, for instance, W L. Ferrar, "Algebra," pp 138–141, Oxford Book Company, Inc , New York, 1941

which, being a sum of squares, can vanish only if

$$x_1 + 2x_2 - 2x_3 = 0 \qquad \text{and} \qquad x_2 = 0 \qquad \text{and} \qquad x_3 = 0$$

and these, in turn, can hold simultaneously only if $x_1 = x_2 = x_3 = 0$.

On the other hand, the quadratic form

$$\|x_1 \quad x_2 \quad x_3\| \left\| \begin{matrix} 1 & 2 & -2 \\ 2 & 3 & -4 \\ -2 & -4 & 5 \end{matrix} \right\| \left\| \begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix} \right\| = \begin{matrix} x_1^2 & + 2x_1x_2 & - 2x_1x_3 \\ +2x_2x_1 & + 3x_2^2 & - 4x_2x_3 \\ -2x_3x_1 & - 4x_3x_2 & + 5x_3^2 \end{matrix}$$

is not definite, since the three quantities

$$1 \qquad \text{and} \qquad \begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix} = -1 \qquad \text{and} \qquad \begin{vmatrix} 1 & 2 & -2 \\ 2 & 3 & -4 \\ -2 & -4 & 5 \end{vmatrix} = -1$$

do not fulfill either of the conditions of Theorem 1. In fact, this quadratic form can be written as $(x_1 + 2x_2 - 2x_3)^2 - x_2^2 + x_3^2$, and, since this expression takes on the value 1 when $x_1 = 2$, $x_2 = 0$, $x_3 = 1$ and takes on the value -1 when $x_1 = -2$, $x_2 = 1$, $x_3 = 0$, it is actually indefinite.

In our definition of a quadratic form, neither the matrix of coefficients A nor the matrix of unknowns X was restricted to be real. However, in most elementary applications both A and X will be real, and only for real quadratic forms are such properties as definiteness and indefiniteness defined. Actually, when complex quantities are involved, quadratic forms, as we have defined them, are almost always replaced by related expressions known as **hermitian forms**:

DEFINITION 1

If A is a hermitian matrix, the expression $\bar{X}^TAX$ is known as a hermitian form.

Recalling the definition of a hermitian matrix (Sec. 10.2) it is easy to verify that any hermitian form is equal to its transposed conjugate. Moreover, since it is a scalar, i.e., a (1,1) matrix, it is also equal to its transpose. Hence, we have the following result:

THEOREM 4

The value of a hermitian form is real for all values of its variables.

Because of Theorem 4, positive- and negative-definite, positive- and negative-semidefinite, and indefinite hermitian forms can be defined precisely as the corresponding types of quadratic forms were defined. Moreover, it can be shown that the criteria for definiteness contained in Theorems 1, 2, and 3 hold without change for hermitian forms.

Closely associated with quadratic forms are what are known as **bilinear forms**:

DEFINITION 2

If A is a symmetric matrix, the expression Y^TAX is known as a bilinear form.

Clearly, if $Y = X$, the bilinear form Y^TAX becomes the quadratic form X^TAX. If the components of Y are thought of as the coordinates of a "point" in a hyperspace of the appropriate number of dimensions, the bilinear form Y^TAX is sometimes called the **polar** of the point Y with respect to the quadratic form X^TAX.

It is interesting to note that the scalar product of two vectors Y and X, namely, Y^TX, can be thought of as the bilinear form Y^TIX The condition that Y and X be orthogonal is, then, just the condition that the bilinear form Y^TIX be equal to zero. This suggests that the simple notion of orthogonality introduced in Sec. 10 2, by analogy with the familiar results of solid analytic geometry, be extended to include the following concept of **generalized orthogonality**:

DEFINITION 3

Two vectors X and Y are said to be orthogonal with respect to a symmetric matrix A if the bilinear form Y^TAX is equal to zero

In the spirit of Definition 3, the notion of the length of a vector can also be generalized In fact, the definition of the length of a vector X introduced in Sec 10 2, namely, $\sqrt{X^TX}$, can be rewritten $\sqrt{X^TIX}$, and this suggests

$$\sqrt{X^TAX}$$

as the **generalized length** of the vector X with respect to the symmetric matrix A This is meaningful, however, only if the quantity under the radical is positive Hence, it is necessary to require further that A be the matrix of a positive-definite quadratic form. A vector whose generalized length with respect to a given symmetric positive-definite matrix is 1 is said to be **normalized** with respect to that matrix A vector X can always be normalized with respect to a given symmetric positive-definite matrix A by dividing it by the positive quantity $\sqrt{X^TAX}$

Clearly, the Schmidt orthogonalization process, which we discussed at the end of Sec 10.5, can be carried out equally well using the concepts of generalized orthogonality and generalized length The notion of orthogonality with respect to a nonunit matrix will be of considerable importance in the work of this chapter

Just as it was convenient in analytic geometry to be able to remove the cross-product term from the equation of a conic, so in many applications involving quadratic forms it is desirable to be able to remove the cross-product terms by a suitable transformation and express the quadratic form as a sum of squares There are many ways of doing this, among which the following, due to Lagrange, is particularly effective. The general idea is first to group together all terms containing x_1 as a factor and, by suitable manipulations, make this expression a perfect square Then, among the terms that remain, those which contain x_2 as a factor are rearranged into an expression which is a perfect square; and so on, until the process terminates The original form will then appear as a sum of squares of linear polynomials in $(x_1,x_2, \ldots ,x_n)$, $(x_2,x_3, \ldots ,x_n)$, . . . , (x_{n-1},x_n), and x_n, respectively, and when these expressions are taken as new variables the reduction is complete.

To begin the process let us assume that $a_{11} \neq 0$ and group together all terms containing x_1 as a factor:

$$\begin{aligned}&(a_{11}x_1^2 + 2a_{12}x_1x_2 + \cdots + 2a_{1n}x_1x_n) + \sum_{r,s=2}^{n} a_{rs}x_rx_s \\ &= a_{11}\left(x_1^2 + \frac{2a_{12}}{a_{11}}x_1x_2 + \cdots + \frac{2a_{1n}}{a_{11}}x_1x_n\right) + \phi_1(x_2,x_3, \ldots ,x_n)\end{aligned}$$

Now, adding and subtracting the appropriate terms, none of which involves x_1, we have

$$\begin{aligned}a_{11}\Bigg[&\left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \cdots + \frac{a_{1n}}{a_{11}}x_n\right)^2 \\ &- \left(\frac{a_{12}}{a_{11}}\right)^2 x_2^2 - \cdots - \left(\frac{a_{1n}}{a_{11}}\right)^2 x_n^2 \\ &- 2\frac{a_{12}}{a_{11}}\frac{a_{13}}{a_{11}}x_2x_3 - \cdots - 2\frac{a_{1,n-1}}{a_{11}}\frac{a_{1n}}{a_{11}}x_{n-1}x_n\Bigg] \\ &\qquad + \phi_1(x_2,x_3, \ldots ,x_n) \\ = a_{11}&\left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \cdots + \frac{a_{1n}}{a_{11}}x_n\right)^2 + \phi_2(x_2,x_3, \ldots ,x_n)\end{aligned}$$

The obviously nonsingular transformation

$$T_1 \qquad \begin{aligned} y_1 &= x_1 + \frac{a_{12}}{a_{11}}x_2 + \cdots + \frac{a_{1n}}{a_{11}}x_n \\ y_2 &= x_2 \\ &\ \ \vdots \\ y_n &= x_n \end{aligned}$$

now reduces $Q(x)$ to the form

$$a_{11}y_1^2 + \phi_2(y_2,y_3, \ldots ,y_n)$$

where $\phi_2(y_2,y_3, \ldots ,y_n)$ is, of course, a quadratic form in the $n - 1$ variables $y_2, y_3, \ldots , y_n$, with coefficients b_{ij}, say

The same process is next applied to ϕ_2, and a second nonsingular transformation, of the form

$$T_2{:} \qquad \begin{aligned} z_1 &= y_1 \\ z_2 &= y_2 + \frac{b_{23}}{b_{22}}y_3 + \cdots + \frac{b_{2n}}{b_{22}}y_n \\ z_3 &= y_3 \\ &\ \ \cdots \\ z_n &= y_n \end{aligned}$$

extends the reduction to

$$a_{11}z_1^2 + b_{22}z_2^2 + \phi_3(z_3,z_4, \ . \ . \ . \ ,z_n)$$

The continuation is now obvious, and the required transformation is, finally, the product of the successive transformations T_1, T_2, . . . , T_n.

If at any stage all square terms are missing from the form $\phi_i(u_i,u_{i+1}, \ldots ,u_n)$ the process must be modified If this occurs, either no more terms remain and the reduction is complete, or else there is at least one cross-product term with nonzero coefficient, say u_ju_{j+1} If this is the case, the nonsingular transformation

$$\begin{aligned} u_1 &= u_1' \\ &\cdot \ \cdot \\ u_{j-1} &= u_{j-1}' \\ u_j &= u_j' + u_{j+1}' \\ u_{j+1} &= u_j' - u_{j+1}' \\ u_{j+2} &= u_{j+2}' \\ &\cdot \ \cdot \ \cdot \ \cdot \\ u_n &= u_n' \end{aligned}$$

will clearly introduce a term in $(u_j')^2$, and the process can be continued in its original form.

It is important to note that the linear transformation employed at each stage is rank-preserving Hence, since the rank of a diagonal matrix is equal to the number of its nonzero diagonal elements, it follows that, when X^TAX is transformed to a sum of squares by the Lagrange reduction, the number of square terms present in the final result is equal to the rank of the matrix of the original form It is also clear that, when a positive-definite quadratic form is reduced to a sum of squares by the Lagrange reduction, the final result must consist of the square of each variable with a positive coefficient

EXAMPLE 2

Find a transformation which will reduce to a sum of squares the quadratic form X^TAX, where

$$A = \left\| \begin{matrix} 1 & -1 & 2 & 0 \\ -1 & 2 & -2 & 1 \\ 2 & -2 & 5 & 1 \\ 0 & 1 & 1 & 4 \end{matrix} \right\|$$

Following the Lagrange procedure, we first group together the terms containing x_1 as a factor, and then complete the square on these terms

$$\begin{aligned} &(x_1^2 - 2x_1x_2 + 4x_1x_3) + (2x_2^2 - 4x_2x_3 + 2x_2x_4 + 5x_3^2 + 2x_3x_4 + 4x_4^2) \\ &\quad = [(x_1 - x_2 + 2x_3)^2 - x_2^2 + 4x_2x_3 - 4x_3^2] \\ &\qquad\qquad + (2x_2^2 - 4x_2x_3 + 2x_2x_4 + 5x_3^2 + 2x_3x_4 + 4x_4^2) \\ &\quad = (x_1 - x_2 + 2x_3)^2 + (x_2^2 + 2x_2x_4 + x_3^2 + 2x_3x_4 + 4x_4^2) \end{aligned}$$

Now we apply the transformation

$$T_1 \qquad \begin{aligned} y_1 &= x_1 - x_2 + 2x_3 \\ y_2 &= \phantom{x_1 - {}} x_2 \\ y_3 &= x_3 \\ y_4 &= x_4 \end{aligned}$$

getting $$y_1^2 + (y_2^2 + 2y_2y_4 + y_3^2 + 2y_3y_4 + 4y_4^2)$$

We next apply the same procedure to the function of y_2, y_3, y_4 which remains

$$\begin{aligned} y_1^2 + (y_2^2 + 2y_2y_4) + (y_3^2 + 2y_3y_4 + 4y_4^2) &= y_1^2 + [(y_2 + y_4)^2 - y_4^2] + (y_3^2 + 2y_3y_4 + 4y_4^2) \\ &= y_1^2 + (y_2 + y_4)^2 + (y_3^2 + 2y_3y_4 + 3y_4^2) \end{aligned}$$

We now apply the transformation

$$T_2 \qquad \begin{aligned} z_1 &= y_1 \\ z_2 &= y_2 + y_4 \\ z_3 &= y_3 \\ z_4 &= y_4 \end{aligned}$$

getting $$z_1^2 + z_2^2 + (z_3^2 + 2z_3z_4 + 3z_4^2)$$

A repetition of the process now yields

$$z_1^2 + z_2^2 + (z_3 + z_4)^2 + 2z_4^2$$

Hence, the final transformation

$$T_3 \qquad \begin{aligned} w_1 &= z_1 \\ w_2 &= z_2 \\ w_3 &= z_3 + z_4 \\ w_4 &= z_4 \end{aligned}$$

reduces the original quadratic form to the expression

$$w_1^2 + w_2^2 + w_3^2 + 2w_4^2 \qquad \text{as required}$$

The single transformation which accomplishes the reduction is of course the product of the transformations T_1, T_2, and T_3, that is, the transformation which results when the y's and z's are eliminated and the w's are expressed directly in terms of the x's. This is easily found to be

$$T \qquad W = PX \qquad \text{where } P = \left\| \begin{matrix} 1 & -1 & 2 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{matrix} \right\|$$

To verify that this transformation actually reduces X^TAX to a sum of squares, it is necessary that the transformation T be solved for X, so that we can substitute for X in the expression X^TAX. To do this, we multiply both sides of the equation $W = PX$ by the inverse of P, which surely exists since P is nonsingular. This gives us

$$T^{-1}: \qquad X = P^{-1}W \qquad \text{where } P^{-1} = \left\| \begin{matrix} 1 & 1 & -2 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{matrix} \right\|$$

Under T^{-1} the original form becomes

$$(P^{-1}W)^TA(P^{-1}W) = W^T(P^{-1})^TAP^{-1}W = W^T[(P^{-1})^TAP^{-1}]W$$

and it is easy to verify that $(P^{-1})^TAP^{-1}$ is indeed the diagonal matrix

$$\begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{Vmatrix}$$

If $f(x_1,x_2, \dots ,x_n)$ is a function of n variables which possesses first partial derivatives in the neighborhood of some "point" P: $(a_1,a_2, \dots ,a_n)$, it is shown in calculus that a necessary condition for f to have a maximum or a minimum at the point P is that at P each of the first partial derivatives of f be zero. In elementary calculus, sufficient conditions for a point P to be a maximum or a minimum of f are usually not obtained, but, with the fundamental properties of quadratic forms available, these can be formulated in a relatively simple way. In doing this, we shall use the Taylor expansion of f around the point P: $(a_1,a_2, \dots ,a_n)$, which means that our conclusions are valid only for functions possessing such expansions.*

Under our assumption that f has a Taylor expansion around the point P: $(a_1,a_2, \dots ,a_n)$, we can write, using the operational notation developed in calculus,

$$\begin{aligned} f(x_1, \dots ,x_n) &= f(a_1, \dots ,a_n) \\ &+ \left[(x_1 - a_1)\frac{\partial}{\partial x_1} + \cdots + (x_n - a_n)\frac{\partial}{\partial x_n} \right] f(x_1, \dots ,x_n) \Big|_{a_1, \dots ,a_n} \\ &+ \frac{1}{2!}\left[(x_1 - a_1)\frac{\partial}{\partial x_1} + \cdots + (x_n - a_n)\frac{\partial}{\partial x_n} \right]^2 f(x_1, \dots ,x_n) \Big|_{a_1, \dots ,a_n} \\ &+ \frac{1}{3!}\left[(x_1 - a_1)\frac{\partial}{\partial x_1} + \cdots + (x_n - a_n)\frac{\partial}{\partial x_n} \right]^3 f(x_1, \dots ,x_n) \Big|_{a_1, \dots ,a_n} \\ &+ \cdots \end{aligned}$$

Now, by hypothesis,

$$\frac{\partial f}{\partial x_1}\Big|_{a_1, \dots ,a_n} = \cdots = \frac{\partial f}{\partial x_n}\Big|_{a_1, \dots ,a_n} = 0$$

Hence, letting $\lambda_i = x_i - a_i$, we have

$$\begin{aligned} (3) \qquad & f(x_1, \dots ,x_n) - f(a_1, \dots ,a_n) \\ &= \frac{1}{2}\left[\lambda_1 \frac{\partial}{\partial x_1} + \cdots + \lambda_n \frac{\partial}{\partial x_n} \right]^2 f(x_1, \dots ,x_n) \Big|_{a_1, \dots ,a_n} \\ &\qquad + \text{terms involving the third and higher} \\ &\qquad \text{powers of the infinitesimals } \lambda_1, \dots , \lambda_n \end{aligned}$$

---

* Actually, if we use Taylor's theorem rather than Taylor's expansion, we need assume only the existence of the third derivatives of f.

Clearly, in the neighborhood of P: $(a_1, \dots, a_n)$ the principal part of the right-hand side in the last expression is in general the first group of terms, which together constitute a quadratic form in the λ's in which, specifically, the coefficient of the product $\lambda_i\lambda_j$ is

$$\left.\frac{\partial^2 f}{\partial x_i\,\partial x_j}\right|_{a_1,\dots,a_n} \quad i \neq j \qquad \text{and} \qquad \left.\frac{1}{2}\frac{\partial^2 f}{\partial x_i^2}\right|_{a_1,\dots,a_n} \quad i = j$$

Now P: $(a_1, \dots, a_n)$ will be a local maximum of f if and only if the difference $f(x_1, \dots, x_n) - f(a_1, \dots, a_n)$ is negative for all sufficiently small values of $\lambda_i = x_i - a_i$ $(i = 1, 2, \dots, n)$ which are not all zero And this will be the case if the quadratic form in the λ's is negative-definite Similarly, P will be a local minimum if and only if the difference $f(x_1, \dots, x_n) - f(a_1, \dots, a_n)$ is positive for all sufficiently small values of λ_i which are not all zero, and this will be the case if the quadratic form in the λ's is positive-definite The point P will be neither a maximum nor a minimum if the difference $f(x_1, \dots, x_n) - f(a_1, \dots, a_n)$ is sometimes positive and sometimes negative in the neighborhood of P, and this will be the case if the quadratic form in the λ's is indefinite Finally, if the quadratic form in the λ's is semidefinite, there are points distinct from, but arbitrarily close to, P at which the form is zero, and it is therefore not the principal part of the right-hand side of Eq (3) In this case, a decision requires a consideration of the cubic (quartic, $\dots$) terms in λ, and takes us beyond the bounds of matrix theory

EXAMPLE 3

Examine the function

$$f(x_1,x_2,x_3) = 35 - 6x_1 + 2x_3 + x_1^2 - 2x_1x_2 + 2x_2^2 + 2x_2x_3 + 3x_3^2$$

for maxima and minima

To determine at what points, if any, the given function may have maxima or minima, we investigate the solutions, if any, of the three equations

$$\begin{aligned}
\frac{\partial f}{\partial x_1} &= -6 + 2x_1 - 2x_2 &&= 0\\
\frac{\partial f}{\partial x_2} &= - 2x_1 + 4x_2 + 2x &&= 0\\
\frac{\partial f}{\partial x_3} &= 2 + 2x_2 + 6x_3 &&= 0
\end{aligned}$$

From these we find that the only possibility for a local extremum is the point P $(8,5,-2)$

Clearly, $f(x_1,x_2,x_3) = 9$ and $\dfrac{\partial f}{\partial x_1} = \dfrac{\partial f}{\partial x_2} = \dfrac{\partial f}{\partial x_3} = 0$ at P Moreover, at P,

$$\frac{\partial^2 f}{\partial x_1^2} = 2 \qquad \frac{\partial^2 f}{\partial x_1\,\partial x_2} = -2 \qquad \frac{\partial^2 f}{\partial x_1\,\partial x_3} = 0$$

$$\frac{\partial^2 f}{\partial x_2^2} = 4 \qquad \frac{\partial^2 f}{\partial x_2\,\partial x_3} = 2 \qquad \frac{\partial^2 f}{\partial x_3^2} = 6$$

Hence,

$$f(x_1,x_2,x_3) - 9 = \frac{1}{2!}[2(x_1 - 8)^2 - 4(x_1 - 8)(x_2 - 5)$$
$$+ 4(x_2 - 5)^2 + 4(x_2 - 5)(x_3 + 2) + 6(x_3 + 2)^2] + $$

The second-degree terms in $(x_1 - 8)$, $(x_2 - 5)$, and $(x_3 + 2)$ in this expansion constitute a quadratic form whose matrix is

$$\left\| \begin{matrix} 1 & -1 & 0 \\ -1 & 2 & 1 \\ 0 & 1 & 3 \end{matrix} \right\|$$

By Theorem 1, this quadratic form is positive-definite, hence, the point $(8,5,-2)$ is a local minimum of the given function

EXERCISES

1 Classify each of the following quadratic forms

a $x_1^2 + 4x_2^2 + 4x_3^2 + 4x_1x_2 + 4x_1x_3 + 6x_2x_3$
b $3x_1^2 + 3x_2^2 + 6x_3^2 - 2x_1x_2 - 4x_1x_3$
c $-x_1^2 - 3x_2^2 - 5x_3^2 + 2x_1x_2 + 2x_1x_3 + 2x_2x_3$
d $2x_1^2 + 2x_2^2 + x_3^2 + 2x_1x_3 + 2x_2x_3$

2 Find a transformation which will reduce each of the following quadratic forms to a sum of squares

a $x_1^2 + 5x_2^2 + 2x_3^2 + 4x_1x_2 + 2x_1x_3 + 6x_2x_3$
b $x_1^2 + 5x_2^2 + 5x_3^2 + 2x_4^2 - 2x_1x_2 + 4x_1x_3 + 2x_1x_4 - 6x_2x_4 + 2x_3x_4$
c $x_1x_2 + x_3x_4$ **d** $x_1x_2 + x_3x_4 + x_4x_5 + x_5x_6$

3 If $f(X) = Y^TAX$, show that $f(\lambda X + \mu Y) = \lambda^2X^TAX + 2\lambda\mu X^TAY + \mu^2Y^TAY$

4 If A is symmetric, show that $Y^TAX = X^TAY$

5 Examine the following functions for maxima and minima

a $2x_1^2 + 2x_1x_2 + x_2^2 + 6x_1 + 6x_2 + 3$
b $x_1^2 - 2x_1x_2 + 2x_1x_3 - 4x_2x_3 + 4x_1 - 4x_2 + 4x_3 + 4$
c $-2x_1^2 - 2x_2^2 - x_3^2 + 2x_1x_2 + 2x_2x_3 + 4x_1 - 4x_2 + 2x_3 - 3$
d $x_1^3 + x_2^3 - 3x_1$ **e** $x_1^3 - 3x_1x_2 + x_2^3$
f $\sin x_1 + \sin x_2 + \cos (x_1 + x_2)$

6 Show that in the neighborhood of any zero of an indefinite quadratic form the form takes on both positive and negative values

7 Prove that a nonsingular quadratic form cannot be semidefinite

8 If X^TAX is a positive-definite quadratic form, show that $(X^TAY)^2 \leqq (X^TAX)(Y^TAY)$, the equality sign holding if and only if either X or Y is a null vector or $X = Y$ (Hint Use the result of Exercise 3)

9 From the vectors $V_1 = \|1 \quad 0 \quad 0\|$, $V_2 = \|0 \quad 1 \quad 0\|$, $V_3 = \|0 \quad 0 \quad 1\|$, construct a set of vectors orthonormal with respect to the matrix

$$\left\| \begin{matrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{matrix} \right\|$$

10 **a** Find the potential energy stored in the system shown in Fig 10 1, Sec. 10 3, as a result of the displacements x_1, x_2, and x_3, and show that it is a positive-definite quadratic function of the x's (Hint The work required to stretch a spring of modulus k a distance s is $\frac{1}{2}ks^2$)
b Work part **a** for the system shown in Fig. 10.2, Sec 10.3

11.2

The characteristic equation of a matrix

In studying linear transformations of the form

$$Y = AX \qquad \text{where } Y = \left\| \begin{matrix} y_1 \\ y_2 \\ \cdot \\ y_n \end{matrix} \right\| \qquad A = \left\| \begin{matrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{matrix} \right\| \qquad X = \left\| \begin{matrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{matrix} \right\|$$

it is an interesting and important problem to determine what vectors, if any, are left unchanged in direction. Since two nontrivial vectors have the same direction if and only if one is a nonzero scalar multiple of the other, this is equivalent to the question of determining those vectors X whose images Y are of the form $Y = \lambda X$, that is, those vectors X such that

$$AX = \lambda X \qquad \text{or} \qquad (A - \lambda I)X = O$$

Clearly, the matric equation $(A - \lambda I)X = O$ is equivalent to the scalar system

$$\begin{aligned} (a_{11} - \lambda)x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + \cdots + a_{2n}x_n &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + (a_{nn} - \lambda)x_n &= 0 \end{aligned}$$

and, according to Corollary 1, Theorem 7, Sec. 10.5, a homogeneous equation of this sort will have one or more nontrivial solutions if and only if the determinant of the coefficients is equal to zero. This condition, namely,

$$(1) \qquad |A - \lambda I| = \left| \begin{matrix} (a_{11} - \lambda) & a_{12} & \cdots & a_{1n} \\ a_{21} & (a_{22} - \lambda) & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & \cdots & \cdots & (a_{nn} - \lambda) \end{matrix} \right| = 0$$

is obviously a polynomial equation of degree n in the parameter λ with leading coefficient $(-1)^n$.

$$(2) \qquad |A - \lambda I| = (-1)^n[\lambda^n - \beta_1\lambda^{n-1} + \beta_2\lambda^{n-2} + (-1)^{n-1}\beta_{n-1}\lambda + (-1)^n\beta_n] = 0$$

Both this equation and the equivalent equation obtained by dropping the factor $(-1)^n$ are known as the **characteristic equation** of the matrix A, and the expression in brackets is known as the **characteristic polynomial** of A. For values of λ which satisfy Eq. (2) and for these values only, the matric equation $(A - \lambda I)X = O$ has nontrivial solution vectors. The n roots of Eq. (2), which of course need not be distinct, are called the **characteristic roots** or **characteristics values** of the matrix A, and the corresponding solutions are called the **characteristic**

vectors* of A. Although we have introduced the idea of characteristic values and characteristic vectors of a matrix in connection with the specific problem of determining the vectors left invariant in direction by a linear transformation, these concepts are of fundamental importance in much of matrix theory and in many of its physical applications.

Since most applications involve matrices which are either real and symmetric or hermitian, we shall for the most part restrict the rest of our discussion to the characteristic values and characteristic vectors of such matrices. We begin, however, with several theorems which deal with the characteristic values and characteristic vectors of arbitrary square matrices.

Since the characteristic equation (2) of a square matrix A is a polynomial equation, its roots, say $\lambda_1, \lambda_2, \ldots, \lambda_n$, are connected with its coefficients $-\beta_1, \beta_2, \ldots, (-1)^n\beta_n$ by the well-known root-coefficient relations

$$\begin{array}{l} \beta_1 = \lambda_1 + \lambda_2 + \cdots + \lambda_n \\ \beta_2 = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \cdots + \lambda_{n-1}\lambda_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \beta_n = \lambda_1\lambda_2 \cdots \lambda_n \end{array} \tag{3}$$

Furthermore, if we set $\lambda = 0$ in Eq. (2), we obtain

$$|A| = (-1)^{2n}\beta_n = \beta_n \tag{4}$$

Hence, from the last of Eqs. (3), we have

$$|A| = \lambda_1\lambda_2 \cdots \lambda_n$$

From this it follows that $|A|$ is zero if and only if at least one of the λ's is zero. Thus we have established the following theorem.

THEOREM 1

A matrix is singular if and only if at least one of its characteristic values is zero.

Equation (4) is only the first of a series of relations connecting the coefficients in Eq. (2) with the principal minors of A. For instance, if $|A - \lambda I|$ is written as the sum of 2^n determinants by repeated use of the addition theorem (Theorem 9, Sec. 10.1), it is clear that the terms containing the first power of λ are obtained by multiplying the term $-\lambda$ in each diagonal element of $|A - \lambda I|$ by the λ-free part of the cofactor of that element. Thus, the coefficient of λ in Eq. (2), namely,

$$(-1)^{2n-1}\beta_{n-1} = -\beta_{n-1}$$

is equal to $\qquad -(A_{11} + A_{22} + \cdots + A_{nn})$

---

* Some writers graft the German word *eigen* meaning *own, peculiar,* or *proper,* onto the words *values* and *vectors* and use the hybrid terms *eigenvalues* and *eigenvectors*.

Hence, it follows that

$$\beta_{n-1} = A_{11} + A_{22} + \cdots + A_{nn}$$

Similarly, the terms containing λ^2 in the expansion of $|A - \lambda I|$ are found by multiplying the terms containing $-\lambda$ in every pair of diagonal elements by the λ-free part of the algebraic complement of the second-order minor containing those diagonal elements. Thus the coefficient of λ^2 in Eq. (2), namely,

$$(-1)^{2n-2}\beta_{n-2} = \beta_{n-2}$$

is equal to $\quad A_{12,12} + A_{13,13} + \cdots + A_{n-1\,n,\,n-1\,n}$

The continuation is obvious, and we, therefore, have the following theorem.

THEOREM 2

If $\lambda^n - \beta_1\lambda^{n-1} + \cdots + (-1)^{n-1}\beta_{n-1}\lambda + (-1)^n\beta_n = 0$ is the characteristic equation of a square matrix A, then β_i is equal to the sum of all the principal minors of order i in A.

For $i = 1$ we have, as a special case of Theorem 2, the relation

$$\beta_1 = \lambda_1 + \lambda_2 + \cdots + \lambda_n = a_{11} + a_{22} + \cdots + a_{nn}$$

The quantity $a_{11} + a_{22} + \cdots + a_{nn}$ is called the **trace** of A.

The characteristic polynomial of a matrix A and, hence, the coefficients $\{\beta_i\}$ and the characteristic roots $\{\lambda_i\}$ have the interesting property of being invariant under any similarity transformation. More precisely, we have the following theorem:

THEOREM 3

If A and B are similar square matrices, then A and B have the same characteristic polynomial.

PROOF Let $|A - \lambda I|$ be the characteristic polynomial of the matrix A, and let B be a matrix similar to A, i.e., let B be any matrix such that $B = S^{-1}AS$. Then the characteristic polynomial of B is

$$\begin{aligned}
|B - \lambda I| &= |S^{-1}AS - \lambda I| \\
&= |S^{-1}AS - \lambda S^{-1}IS| \\
&= |S^{-1}(A - \lambda I)S| \\
&= |S^{-1}| \cdot |A - \lambda I| \cdot |S|
\end{aligned}$$

since the determinant of a product of square matrices is equal to the product of the determinants of the individual matrices. Moreover, by Corollary 2, Theorem 1, Sec. 10.3, $|S^{-1}| \cdot |S| = 1$. Hence, $|B - \lambda I| = |A - \lambda I|$, as asserted.

The next three theorems also deal with the characteristic values and characteristic vectors of arbitrary square matrices:

THEOREM 4

A characteristic vector of a square matrix cannot correspond to two distinct characteristic values.

PROOF Let λ_1 and λ_2 be distinct characteristic values of a square matrix A, and let X_1 be a characteristic vector of A corresponding, if possible, to both λ_1 and λ_2 Then, simultaneously,

$$(A - \lambda_1 I)X_1 = O \qquad \text{and} \qquad (A - \lambda_2 I)X_1 = O$$

Hence, subtracting,

$$(\lambda_2 - \lambda_1)IX_1 = (\lambda_2 - \lambda_1)X_1 = O \tag{5}$$

However, by hypothesis, $\lambda_1 \neq \lambda_2$ Moreover, a characteristic vector is, by definition, a nontrivial solution vector of $(A - \lambda I)X = O$ Thus $X_1 \neq O$, and therefore Eq. (5) cannot hold Hence, the assumption that a characteristic vector can correspond to two distinct characteristic values must be abandoned, and the theorem is established.

THEOREM 5

If $X_1, X_2, \ldots, X_m$ $(m \leqq n)$ are characteristic vectors corresponding respectively to the distinct characteristic values $\lambda_1, \lambda_2, \ldots, \lambda_m$ of an (n,n) matrix A, then $X_1, X_2, \ldots, X_m$ are linearly independent

PROOF Let $X_1, X_2, \ldots, X_m$ be characteristic vectors corresponding respectively to the distinct characteristic values $\lambda_1, \lambda_2, \ldots, \lambda_m$ of a square matrix A, and let us suppose, contrary to the theorem, that $X_1, X_2, \ldots, X_m$ are dependent More specifically, let us suppose that the maximum number of linearly independent vectors in the set is k, where $1 \leqq k < m$, and, for convenience, let them be the first k X's Then the relation

$$\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_k X_k = O$$

implies that $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$, but there does exist a nontrivial set of β's, with $\beta_{k+1} \neq 0$, such that

$$\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \beta_{k+1} X_{k+1} = O \tag{6}$$

Now multiply Eq. (6) on the left by the matrix A, getting

$$\beta_1 AX_1 + \beta_2 AX_2 + \cdots + \beta_k AX_k + \beta_{k+1} AX_{k+1} = O$$

However, $AX_i = \lambda_i X_i$ for each i Hence the last equation becomes

$$\beta_1\lambda_1 X_1 + \beta_2\lambda_2 X_2 + \cdots + \beta_k\lambda_k X_k + \beta_{k+1}\lambda_{k+1} X_{k+1} = O \tag{7}$$

If we now multiply Eq (6) by λ_{k+1} and subtract from Eq (7), we obtain

$$(\lambda_1 - \lambda_{k+1})\beta_1 X_1 + (\lambda_2 - \lambda_{k+1})\beta_2 X_2 + \cdots + (\lambda_k - \lambda_{k+1})\beta_k X_k = O \tag{8}$$

Since $X_1, X_2, \ldots, X_k$ are linearly independent, by hypothesis, it follows that each coefficient in (8) is equal to zero. Hence, since $\lambda_i - \lambda_{k+1} \neq 0$ $(i = 1, 2, \ldots, k)$, it must be that

$$\beta_i = 0 \qquad i = 1, 2, \ldots, k$$

But, if this is the case, it follows from (6) that

$$\beta_{k+1} X_{k+1} = O$$

which is impossible, since neither the scalar β_{k+1} nor the vector X_{k+1} is zero This contradiction overthrows the possibility that the characteristic vectors $X_1, X_2, \ldots, X_m$ are linearly dependent, and the theorem is established.

In particular, if an (n,n) matrix has n distinct characteristic values, the last theorem tells us that it has n linearly independent characteristic vectors Hence, using Corollary 1, Theorem 10, Sec 10.5, we have the following result·

COROLLARY 1

If the characteristic values of an (n,n) matrix A are all distinct, then A has n linearly independent characteristic vectors, and any vector with n components can be expressed as a linear combination of the characteristic vectors of A.

Since the characteristic equation of an (n,n) matrix is always of degree n, it is obvious that, if repeated roots are counted the appropriate number of times, such a matrix always has exactly n characteristic roots With the same convention one might perhaps be able to say that an (n,n) matrix always has exactly n characteristic vectors However, attempting to assign a multiplicity to a characteristic vector associated with a repeated characteristic root is completely artificial and without significance The decisive consideration is the number of linearly independent characteristic vectors of a given matrix, hence, it is of fundamental importance to know when more than one independent characteristic vector is associated with a repeated characteristic root The next theorem gives us a partial answer to this question·

THEOREM 6

If λ_1 is a characteristic root of multiplicity r of an (n,n) matrix A, then the rank of $A - \lambda_1 I$ is equal to or greater than $n - r$

PROOF If A is an (n,n) matrix and if $\lambda = \lambda_1$ is a repeated root of multiplicity r of the characteristic equation $|A - \lambda I| = 0$, then, writing $\lambda = \lambda_1 + w$, it is clear that $w_1 = 0$ is a repeated root of multiplicity r of the equation

$$|A - (\lambda_1 + w)I| = |(A - \lambda_1 I) - wI| = 0$$

Hence, the expanded form of the last equation, say

$$(-1)^n[w^n - \sigma_1 w^{n-1} + \sigma_2 w^{n-2} - \qquad + (-1)^{n-1}\sigma_{n-1}w + (-1)^n\sigma_n] = 0$$

must contain w^r as a factor and must, therefore, reduce to

$$(-1)^n[w^n - \sigma_1 w^{n-1} + \qquad + (-1)^{n-r}\sigma_{n-r}w^r] = 0$$

where $\sigma_{n-r} \neq 0$ Now, by Theorem 2, the coefficient σ_{n-r} is equal to the sum of all principal minors of order $n - r$ of the matrix $A - \lambda_1 I$ Hence, since $\sigma_{n-r} \neq 0$, at least one of these minors must be different from zero In other words, the rank of $A - \lambda_1 I$ must be at least as great as $n - r$, as asserted

If, for a particular root λ_1 of multiplicity r of an (n,n) matrix A, the equality sign holds in the assertion of Theorem 6, then, according to Theorem 6, Sec. 10 5, there are exactly $n - (n - r) = r$ linearly independent characteristic vectors associated with λ_1 Such a characteristic root is said to be **regular.** However, this is the exception rather than the rule, and, in general, there will be a single independent characteristic vector associated with a repeated characteristic root of any multiplicity. For instance, for the matrix

$$A = \left\|\begin{matrix} -3 & -7 & -5 \\ 2 & 4 & 3 \\ 1 & 2 & 2 \end{matrix}\right\|$$

we have
$$|A - \lambda I| = \begin{vmatrix} -3-\lambda & -7 & -5 \\ 2 & 4-\lambda & 3 \\ 1 & 2 & 2-\lambda \end{vmatrix}$$
$$= -\lambda^3 + 3\lambda^2 - 3\lambda + 1 = -(\lambda - 1)^3 = 0$$

Thus, A has a single characteristic root $\lambda = 1$ Moreover, for $\lambda = 1$, the rank of

$$|A - \lambda I|_{\lambda=1} = |A - I| = \begin{vmatrix} -4 & -7 & -5 \\ 2 & 3 & 3 \\ 1 & 2 & 1 \end{vmatrix}$$

is clearly 2 Hence, according to Theorem 8, Sec 10 5, the system of equations $(A - I)X = O$ has a single independent solution vector, namely,

$$X = \left\|\begin{matrix} -3 \\ 1 \\ 1 \end{matrix}\right\|$$

and A has just one independent characteristic vector

Later in this section we shall see that for hermitian matrices the assertion of the last theorem can be sharpened to a strict equality, in other words, we shall prove that, if $\lambda = \lambda_1$ is a characteristic root of multiplicity r of a hermitian matrix A, then the rank of $|A - \lambda_1 I|$ is exactly $n - r$ Preparatory to this, however, it will be convenient to prove first some other theorems about hermitian matrices

THEOREM 7

The characteristic values of a hermitian matrix are all real

PROOF Let A be a hermitian matrix; let λ_1 be any one of its characteristic values; and let X_1 be a characteristic vector corresponding to λ_1 Then

$$(A - \lambda_1 I)X_1 = O$$

or

$$AX_1 = \lambda_1 X_1 \tag{9}$$

and from this, by premultiplying by $\bar{X}_1^T$, we obtain

$$\bar{X}_1^T A X_1 = \lambda_1 \bar{X}_1^T X_1 \tag{10}$$

Now, from the properties of conjugate complex numbers, $\bar{X}_1^T X_1$ is real and in fact positive Furthermore, from Theorem 4, Sec 11.1, we know that $\bar{X}_1^T A X_1$ is also real. Hence, it follows immediately from Eq. (10) that λ_1 is real, as asserted.

Since, as we observed in Sec. 10.2, a real symmetric matrix is just a special case of a hermitian matrix, we have the following important corollary of Theorem 7

COROLLARY 1

The characteristic values of a real symmetric matrix are all real

Furthermore, since iA is hermitian if A is skew-hermitian and since $|A - \lambda I| = 0$ implies $|iA - i\lambda I| = 0$, it follows that, if λ_1 is a characteristic value of the skew-hermitian matrix A, then $i\lambda_1$ is a characteristic value of the hermitian matrix iA. Hence, by Theorem 7, $i\lambda_1$ is real, and therefore, λ_1 is a pure imaginary Thus we have established the following result:

COROLLARY 2

The characteristic values of a skew-hermitian matrix are all pure imaginary

Knowing now that the characteristic roots of a hermitian matrix A are all real, we can return to the characteristic equation of A and prove the following result

THEOREM 8

If $\lambda^n - \beta_1\lambda^{n-1} + \beta_2\lambda^{n-2} - \cdots + (-1)^{n-1}\beta_{n-1}\lambda + (-1)^n\beta_n = 0$ is the characteristic equation of a hermitian matrix A, then the characteristic roots of A are all positive if and only if each β is positive

PROOF If A is a hermitian matrix, it follows from Theorem 7 that the roots of the characteristic equation

$$\lambda^n - \beta_1\lambda^{n-1} + \beta_2\lambda^{n-2} - \cdots + (-1)^{n-1}\beta_{n-1}\lambda + (-1)^n\beta_n = 0$$

are all real Furthermore, if each β is positive, it follows by Descartes's rule of signs that no root of the characteristic equation can be negative or zero Hence, all the characteristic roots must be positive Conversely, if the characteristic roots of A are all positive, then from the root-coefficient relations

$$\begin{aligned}
\beta_1 &= \lambda_1 + \lambda_2 + \cdots + \lambda_n \\
\beta_2 &= \lambda_1\lambda_2 + \lambda_1\lambda_3 + \cdots + \lambda_{n-1}\lambda_n \\
\beta_n &= \lambda_1\lambda_2 \cdots \lambda_n
\end{aligned}$$

it follows at once that each β is positive, as asserted

COROLLARY 1

If $\lambda^n - \beta_1\lambda^{n-1} + \beta_2\lambda^{n-2} - \cdots + (-1)^{n-1}\beta_{n-1}\lambda + (-1)^n\beta_n = 0$ is the characteristic equation of a real symmetric matrix A, then the characteristic roots of A are all positive if and only if each β is positive

One of the most important properties of the characteristic vectors of a hermitian matrix is that of orthogonality. More precisely, we have the following theorem

THEOREM 9

If X_i and X_j are characteristic vectors corresponding, respectively, to the distinct characteristic values λ_i and λ_j of a hermitian matrix A, then $\bar{X}_i^T X_j = 0$.

PROOF By hypothesis, we have

$$AX_i = \lambda_i X_i \tag{11}$$

$$AX_j = \lambda_j X_j \tag{12}$$

If in the first of these we take the conjugate and then the transpose of each member, we obtain

$$\bar{X}_i^T \bar{A}^T = \bar{\lambda}_i \bar{X}_i^T$$

or, since $\bar{A}^T = A$, by hypothesis, and $\bar{\lambda}_i = \lambda_i$, by Theorem 7,

$$\bar{X}_i^T A = \lambda_i \bar{X}_i^T \tag{13}$$

Now, if we premultiply Eq (12) by $\bar{X}_i^T$ and postmultiply Eq (13) by X_j, we obtain, respectively,

$$\bar{X}_i^T A X_j = \lambda_j \bar{X}_i^T X_j$$

$$\bar{X}_i^T A X_j = \lambda_i \bar{X}_i^T X_j$$

Finally, subtracting these equations, we have

$$(\lambda_i - \lambda_j)\bar{X}_i^T X_j = 0$$

or, since $\lambda_i \neq \lambda_j$ by hypothesis,

$$\bar{X}_i^T X_j = 0$$

as asserted

COROLLARY 1

If X_i and X_j are characteristic vectors corresponding to the distinct characteristic values λ_i and λ_j of a real symmetric matrix, then $X_i^T X_j = 0$

We are now in a position to return to the question we raised earlier in this section about the rank of $|A - \lambda_i I|$ when A is hermitian and λ_i is a characteristic root of multiplicity r As the next theorem shows, every characteristic root of a hermitian matrix is regular, that is, if A is hermitian, then, for every characteristic root λ_i of multiplicity r, the rank of $|A - \lambda_i I|$ drops to the minimum permitted by Theorem 6, namely, $n - r$, and there are r linearly independent characteristic vectors corresponding to λ_i

THEOREM 10

If A is a hermitian matrix, then to every r-fold characteristic root of A there correspond exactly r linearly independent characteristic vectors

PROOF Let λ_1 be a repeated characteristic root of a hermitian matrix A, let U_1 be any normalized characteristic vector corresponding to λ_1, and let U be any unitary matrix having U_1 as its first column In virtue of the Schmidt orthogonalization process, it is clear that such a matrix exists Moreover, since U is unitary, that is, since $\bar{U}^T = U^{-1}$, it follows from Theorem 3 that the matrices

$$\bar{U}^T A U - \lambda I \qquad \text{and} \qquad A - \lambda I$$

have the same characteristic equation and, therefore, the same characteristic roots. Now let us write U in partitioned form.

$$U = \| U_1 \; U_2 \; \cdots \; U_n \|$$

Then, remembering that $AU_1 = \lambda_1 U_1$ since, by hypothesis, U_1 is a characteristic vector of A corresponding to λ_1, we have

$$\bar{U}^T A U = \begin{Vmatrix} \bar{U}_1^T \\ \bar{U}_2^T \\ \\ \bar{U}_n^T \end{Vmatrix} \cdot A \| U_1 \; U_2 \; \cdots \; U_n \| = \begin{Vmatrix} \bar{U}_1^T \\ \bar{U}_2^T \\ \\ \bar{U}_n^T \end{Vmatrix} \| AU_1 \; AU_2 \; \cdots \; AU_n \|$$

$$= \begin{Vmatrix} \bar{U}_1^T \\ \bar{U}_2^T \\ \\ \bar{U}_n^T \end{Vmatrix} \| \lambda_1 U_1 \; AU_2 \; \cdots \; AU_n \| = \begin{Vmatrix} \lambda_1 & \bar{U}_1^T A U_2 & \cdots & \bar{U}_1^T A U_n \\ 0 & \bar{U}_2^T A U_2 & \cdots & \bar{U}_2^T A U_n \\ \cdot & \cdot \; \cdot & \cdot \; \cdot \; \cdot & \cdot \; \cdot \; \cdot \\ 0 & \bar{U}_n^T A U_2 & \cdots & \bar{U}_n^T A U_n \end{Vmatrix}$$

where the zeros in the first column of the last matrix enter because U is a unitary matrix, and, therefore, any two of its columns satisfy the relation

$$\bar{U}_i^T U_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

The remaining entries are not in general zero, since U_i and U_j are not orthogonal with respect to the matrix A. However, since A and, therefore, $\bar{U}^T A U$ are hermitian (see Exercise 8), it follows that the elements after the first in the first row must all be zero. Thus

$$\bar{U}^T A U = \begin{Vmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \alpha_{22} & \alpha_{23} & \cdots & \alpha_{2n} \\ 0 & \alpha_{32} & \alpha_{33} & \cdots & \alpha_{3n} \\ \cdot & \cdot \; \cdot & \cdot & \cdot \; \cdot & \cdot \; \cdot \\ 0 & \alpha_{n2} & \alpha_{n3} & \cdots & \alpha_{nn} \end{Vmatrix}$$

and

$$\bar{U}^T A U - \lambda I = \begin{Vmatrix} \lambda_1 - \lambda & 0 & 0 & \cdots & 0 \\ 0 & \alpha_{22} - \lambda & \alpha_{23} & \cdots & \alpha_{2n} \\ 0 & \alpha_{32} & \alpha_{33} - \lambda & \cdots & \alpha_{3n} \\ \cdot & \cdot \; \cdot & \cdot & \cdot & \cdot \; \cdot \; \cdot \\ 0 & \alpha_{n2} & \alpha_{n3} & \cdots & \alpha_{nn} - \lambda \end{Vmatrix}$$

Therefore, if λ_1 is a repeated root of

$$|\bar{U}^T A U - \lambda I| = |A - \lambda I| = 0$$

then $\lambda_1 - \lambda$ must be a factor of the minor of the element in the first row and first column of $|A - \lambda_1 I|$. But if this minor vanishes when $\lambda = \lambda_1$, then the rank of $A - \lambda_1 I$ is at most $n - 2$, since all other minors of order $n - 1$ obviously contain either a column of zeros or a row of zeros. Hence, by Theorem 6, Sec. 10.5, $|A - \lambda I| X = O$ has at least two linearly independent solution vectors, and A has at least two linearly independent characteristic vectors.

If the multiplicity of λ_1 is more than 2, the preceding argument can be repeated, using this time any unitary matrix U whose first two columns are any two orthonormal characteristic vectors corresponding to λ_1. This leads to the conclusion that $\lambda_1 - \lambda$ must be a factor of the complementary minor of the second-order minor in the first two rows and first two columns of

$$|\bar{U}^T A U - \lambda I| = |A - \lambda I|$$

Hence, since all other $(n - 2)$-order minors obviously vanish, it is evident that, when $\lambda = \lambda_1$, the rank of $A - \lambda I$ is not more than $n - 3$, and A, therefore, has at least three linearly independent characteristic vectors Clearly, this procedure can be continued until we reach the conclusion that, if λ_1 is an r-fold characteristic root of A, then the rank of A is at most $n - r$, and hence A has at least r independent characteristic vectors But, by Theorem 9, A can have at most r independent characteristic vectors corresponding to an r-fold characteristic root. Hence, A must have exactly r linearly independent characteristic vectors, as asserted

Since, as we have repeatedly observed, a real symmetric matrix is a special case of a hermitian matrix, it is clear that we also have the following result

COROLLARY 1

If A is a real symmetric matrix, then to every r-fold characteristic root of A there correspond exactly r linearly independent characteristic vectors

We are now in a position to prove the following fundamental theorem

THEOREM 11

Every (n,n) hermitian matrix has n linearly independent characteristic vectors

PROOF Let A be an (n,n) hermitian matrix. It may, of course, possess one or more repeated characteristic roots, but, if it does, we know from the last theorem that to each root of multiplicity r there correspond exactly r linearly independent characteristic vectors Hence, A cannot have more than n linearly independent characteristic vectors Specifically, let the characteristic roots of A be

$$\lambda_1, \lambda_2, \ldots, \lambda_k \qquad 1 \leqq k \leqq n$$

let the multiplicity of λ_i be r_i, where $\sum_{i=1}^{k} r_i = n$, and let

$$X_{i1}, X_{i2}, \ldots, X_{ir_i}$$

be r_i independent characteristic vectors corresponding to λ_i Suppose, now, contrary to the assertion of the theorem, that these n characteristic vectors of A are not linearly independent. Then there exists a relation of the form

$$(14) \qquad (c_{11}X_{11} + \cdots + c_{1r_1}X_{1r_1}) + (c_{21}X_{21} + \cdots + c_{2r_2}X_{2r_2}) + \cdots + (c_{k1}X_{k1} + \cdots + c_{kr_k}X_{kr_k}) = O$$

in which at least one c is different from zero

Now consider a typical group of terms, say the ith, in the last expression By Theorem 3, Sec 10.5, unless the c's in such a group are all zero, the combination defines a characteristic vector corresponding to the characteristic value $\lambda = \lambda_i$. Thus, Eq (14) is simply an expression of the form

$$c_1X_1 + c_2X_2 + \cdots + c_kX_k$$

in which each c is either 0 or 1 and at least one c is different from zero. But since the X's now correspond to distinct characteristic values, it follows from Theorem 2 that they are linearly independent and, hence, that each c must be zero. This contradiction establishes the theorem

COROLLARY 1

Every real symmetric (n,n) matrix has n linearly independent characteristic vectors

If an (n,n) matrix has n linearly independent characteristic vectors, then, by means of the Schmidt orthogonalization process applied to the vectors in each of the sets corresponding to a repeated root, a set of normalized mutually orthogonal characteristic vectors can always be constructed Hence, we have the following important result

COROLLARY 2

Every (n,n) hermitian or real symmetric matrix has a set of n orthonormal characteristic vectors

An (n,n) matrix whose columns are orthonormal characteristic vectors of an (n,n) matrix A is said to be a **modal matrix** of A.

In many applications in physics, chemistry, and engineering it is necessary to consider matric equations of the form $(A - \lambda B)X = O$ in which A and B are either hermitian or real and symmetric Such an equation will, of course, have nontrivial solutions if and only if the determinant of the coefficients is equal to zero Paralleling our earlier terminology, the equation $|A - \lambda B| = 0$ is called the **characteristic equation** of the system, its roots are called the **characteristic roots** or **characteristic values** of the system, and the corresponding nontrivial solutions are called the **characteristic vectors** of the system. As one should expect, the theory of the equation $(A - \lambda B)X = O$ resembles closely the theory of the equation $(A - \lambda I)X = O$ In particular, we have the following results

THEOREM 12

The equation $(A - \lambda B)X = O$ has zero as a characteristic root if and only if A is singular

PROOF This follows immediately from a consideration of the characteristic equation $|A - \lambda B| = 0$ when the left-hand side is expressed as a polynomial in λ

THEOREM 13

If A and B are hermitian (or real symmetric) matrices and if B is definite, then the characteristic values of $(A - \lambda B)X = O$ are all real

PROOF Let A and B be hermitian (or real symmetric) matrices, let B be definite, and let X_1 be a characteristic vector of the equation

$$(A - \lambda B)X = O$$

corresponding to the characteristic value λ_1 Then

(15) $$AX_1 = \lambda_1 BX_1$$

Hence, premultiplying Eq (15) by $\bar{X}_1^T$, we have

(16) $$\bar{X}_1^T A X_1 = \lambda_1 \bar{X}_1^T B X_1$$

Now, from Theorem 4, Sec 11 1, we know that both $\bar{X}_1^T A X_1$ and $\bar{X}_1^T B X_1$ are real numbers. Moreover, since B is definite, $\bar{X}_1^T B X_1 \neq 0$ Hence it follows from Eq. (16) that λ_1 is real, as asserted

By inspection of Eq (16), the following results are obtained immediately

COROLLARY 1

If A and B are hermitian (or real symmetric) matrices which are both positive-definite or both negative-definite, then the characteristic values of

$$(A - \lambda B)X = O$$

are all positive.

COROLLARY 2

If A and B are hermitian (or real symmetric) matrices and if A is positive-definite and B is negative-definite or vice versa, then the characteristic values of

$$(A - \lambda B)X = O$$

are all negative.

THEOREM 14

If A and B are hermitian (or real symmetric) matrices and if $X_1, X_2, \ldots, X_k$ are characteristic vectors of the equation $(A - \lambda B)X = O$ corresponding, respectively, to the distinct characteristic values $\lambda_1, \lambda_2, \ldots, \lambda_k$, then the X's satisfy the generalized orthogonality condition

$$\bar{X}_i^T B X_j = 0 \qquad (\text{or } X_i^T B X_j = 0) \qquad i \neq j$$

PROOF Let A and B be hermitian matrices, let λ_i and λ_j be distinct characteristic values of the equation $(A - \lambda B)X = O$, and let X_i and X_j be characteristic vectors corresponding respectively to λ_i and λ_j Then

$$AX_i = \lambda_i BX_i \qquad \text{and} \qquad AX_j = \lambda_j BX_j$$

If we premultiply the first of these equations by $\bar{X}_j^T$ and the second by $\bar{X}_i^T$, we obtain, respectively,

(17) $$\bar{X}_j^T A X_i = \lambda_i \bar{X}_j^T B X_i$$

and

(18) $$\bar{X}_i^T A X_j = \lambda_j \bar{X}_i^T B X_j$$

Now, if we take the transpose and then the conjugate of each side of Eq (17), remembering that A and B are hermitian, we obtain

(19) $$\bar{X}_i^T A X_j = \lambda_i \bar{X}_i^T B X_j$$

Finally, subtracting Eq (18) from Eq (19), we have

$$(\lambda_i - \lambda_j)\bar{X}_i^T B X_j = 0$$

Therefore, since $\lambda_i \neq \lambda_j$, by hypothesis, it follows that

$$\bar{X}_i^T B X_j = 0 \qquad i \neq j \qquad \text{as asserted}$$

If A and B are real symmetric matrices, an almost identical proof carried through with X_j^T and X_i^T replacing $\bar{X}_j^T$ and $\bar{X}_i^T$, serves to establish the parenthetical assertion of the theorem

COROLLARY 1

If A and B are hermitian (or real symmetric) matrices, and if X_i and X_j are characteristic vectors of $(A - \lambda B)X = O$ corresponding, respectively, to the distinct characteristic values λ_i and λ_j, then $\bar{X}_i^T A X_j = 0$ (or $X_i^T A X_j = 0$)

PROOF This result follows immediately from Eq (18) (or the equation $X_i^T A X_j = \lambda_j X_i^T B X_j$ if A and B are real symmetric matrices) and the fact, guaranteed by Theorem 14, that $\bar{X}_i^T B X_j = 0$ (or $X_i^T B X_j = 0$)

THEOREM 15

If A and B are hermitian (or real symmetric matrices), if B is either positive-definite or negative-definite, and if $X_1, X_2, \ldots, X_k$ are characteristic vectors of $(A - \lambda B)X = O$ corresponding, respectively, to the distinct characteristic values $\lambda_1, \lambda_2, \ldots, \lambda_k$, then $X_1, X_2, \ldots, X_k$ are linearly independent

PROOF Let A and B be hermitian matrices, let B be definite, and let us suppose, contrary to the theorem, that the characteristic vectors $X_1, X_2, \ldots, X_k$ corresponding respectively to the distinct characteristic values $\lambda_1, \lambda_2, \ldots, \lambda_k$ of $(A - B)X = O$ are linearly dependent Then there exists a relation of the form

$$c_1X_1 + c_2X_2 + \cdots + c_kX_k = O$$

in which at least one of the c's, say c_i, is different from zero Now, if we multiply the last equation through on the left by $\bar{X}_i^T B$,[†] we get

$$c_1\bar{X}_i^T B X_1 + c_2\bar{X}_i^T B X_2 + \cdots + c_i\bar{X}_i^T B X_i + \cdots + c_k\bar{X}_i^T B X_k = 0$$

However, from the orthogonality guaranteed by Theorem 14, it follows that every term in the last equation except $c_i\bar{X}_i^T B X_i$ is equal to zero Moreover, by hypothesis, B is either positive-definite or negative-definite Hence, $\bar{X}_i^T B X_i \neq 0$, and, therefore, $c_i = 0$, contrary to the assumption of linear dependence This contradiction shows that the X's must be linearly independent, and the theorem is established

The last theorem must not be misinterpreted as asserting that if A and B are hermitian (or real symmetric) (n,n) matrices and if B is definite, then $(A - \lambda B)X = O$ has n linearly independent characteristic vectors It guarantees that characteristic vectors corresponding to *distinct* characteristic values of

---

[†] This procedure suffices when A and B are real, symmetric matrices, as well as when they are hermitian because, although Theorem 14 does not assert it explicitly, it is clear that $\bar{X}_i^T B X_j = 0$, $i \neq j$ must also hold for real symmetric matrices, since these are just special cases of hermitian matrices.

$$(A - \lambda B)X = O$$

are linearly independent, but it says nothing about how many distinct characteristic values there are or about how many independent characteristic vectors correspond to a repeated characteristic value. If, because of repeated roots,

$$(A - \lambda B)X = O$$

has fewer than n distinct characteristic values, then, for all we know at present, $(A - \lambda B)X = O$ has fewer than n linearly independent characteristic vectors However, by a proof very much like the proof of Theorem 10, the following result can be established

THEOREM 16

If A and B are hermitian (or real symmetric) matrices and if B is either positive-definite or negative-definite, then to a repeated characteristic value of

$$(A - \lambda B)X = O$$

of multiplicity r there correspond exactly r linearly independent characteristic vectors

With this theorem available, it is not difficult to establish the following counterpart of Theorem 11

THEOREM 17

If A and B are hermitian (or real symmetric) (n,n) matrices and if B is either positive-definite or negative-definite, then the equation $(A - \lambda B)X = O$ has exactly n linearly independent characteristic vectors

By a straightforward application of the Schmidt orthogonalization process applied to the n linearly independent characteristic vectors of $(A - \lambda B)X = O$ guaranteed by Theorem 17, we can establish the following useful results

COROLLARY 1

If A and B are hermitian (or real symmetric) (n,n) matrices and if B is definite, then $(A - \lambda B)X = O$ possesses n characteristic vectors orthogonal with respect to B

COROLLARY 2

If A and B are hermitian (or real symmetric) (n,n) matrices and if B is positive-definite, then $(A - \lambda B)X = O$ possesses n characteristic vectors orthonormal with respect to B

With Theorem 17 and its corollaries available, it is now an easy matter to express an arbitrary vector C with n components as a linear combination of the characteristic vectors of the equation $(A - \lambda B)X = O$, provided A and B are hermitian and B is definite For we can write

$$C = c_1X_1 + c_2X_2 + \cdots + c_nX_n \tag{20}$$

where the X's are characteristic vectors of $(A - \lambda B)X = O$ mutually orthogonal with respect to B. Then, if we premultiply Eq (17) by $\bar{X}_i^TB$ we obtain

$$\bar{X}_i^TBC = c_1\bar{X}_i^TBX_1 + \cdots + c_i\bar{X}_i^TBX_i + \cdots + c_n\bar{X}_i^TBX_n$$

From the orthogonality of the X's, it follows that every term on the right except $c_i\bar{X}_i^TBX_i$ is equal to zero. Moreover, since B is definite, it follows that $\bar{X}_i^TBX_i \neq 0$ Hence, we can solve for c_i, getting

$$c_i = \frac{\bar{X}_i^TBC}{\bar{X}_i^TBX_i} \qquad i = 1, 2, \ldots, n$$

If the X's have been normalized with respect to B, that is, if $\bar{X}_i^TBX_i = 1$ $(i = 1, 2, \ldots, n)$, the last formula reduces to the simpler expression

$$c_i = \bar{X}_i^TBC \qquad i = 1, 2, \ldots, n$$

The fact that we were able to solve for the coefficients in the expression (20) without solving any simultaneous equations should make clear the great convenience of working with a set of vectors which are orthogonal

EXERCISES

1 Find the characteristic values and the corresponding characteristic vectors of each of the following matrices

a $\begin{Vmatrix} 4 & 6 & 6 \\ 1 & 3 & 2 \\ -1 & -5 & -2 \end{Vmatrix}$ **b** $\begin{Vmatrix} 7 & -2 & -4 \\ 3 & 0 & -2 \\ 6 & -2 & -3 \end{Vmatrix}$ **c** $\begin{Vmatrix} 11 & -4 & -7 \\ 7 & -2 & -5 \\ 10 & -4 & -6 \end{Vmatrix}$

d $\begin{Vmatrix} 4 & 6 & 6 \\ 1 & 3 & 2 \\ -1 & -4 & -3 \end{Vmatrix}$ **e** $\begin{Vmatrix} 1 & 1 & 1 \\ 1 & 3 & 3 \\ 2 & 1 & 4 \end{Vmatrix}$ **f** $\begin{Vmatrix} -4 & 5 & 5 \\ -5 & 6 & 5 \\ -5 & 5 & 6 \end{Vmatrix}$

For which of these, if any, are the characteristic vectors orthogonal?

2 Find the characteristic values and the corresponding characteristic vectors for the equation $(A - \lambda B)X = O$, if

a $A = \begin{Vmatrix} 8 & -2 & 0 \\ -2 & 3 & -1 \\ 0 & -1 & 2 \end{Vmatrix}$ $B = \begin{Vmatrix} 8 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{Vmatrix}$

b $A = \begin{Vmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{Vmatrix}$ $B = \begin{Vmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{Vmatrix}$

c $A = \begin{Vmatrix} 6 & -3 & 0 \\ -3 & 6 & -3 \\ 0 & -3 & 4 \end{Vmatrix}$ $B = \begin{Vmatrix} 6 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{Vmatrix}$

d $A = \begin{Vmatrix} 3 & -1 & 0 \\ -1 & 1 & -1 \\ 0 & -1 & 5 \end{Vmatrix}$ $B = \begin{Vmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{Vmatrix}$

In each case, verify all orthogonality relations, and, using orthogonality properties, express the vector $V = \left\|\begin{matrix}1\\2\\3\end{matrix}\right\|$ as a linear combination of the characteristic vectors

3 Find three solution vectors of the equation $(A - \lambda B)X = O$ which are orthonormal with respect to B if

a $A = \left\|\begin{matrix}7 & 1 & -1\\1 & 4 & -1\\-1 & -1 & 3\end{matrix}\right\|$ $B = \left\|\begin{matrix}6 & 0 & 0\\0 & 3 & 0\\0 & 0 & 2\end{matrix}\right\|$

b $A = \left\|\begin{matrix}7 & -1 & -1\\-1 & 4 & 1\\-1 & 1 & 3\end{matrix}\right\|$ $B = \left\|\begin{matrix}6 & 0 & 0\\0 & 3 & 0\\0 & 0 & 2\end{matrix}\right\|$

4 Prove Theorem 17.

5 Prove Corollary 1, Theorem 17

6 In Corollary 2, Theorem 17, why is it necessary to restrict B to being positive-definite?

7 Prove that iA is hermitian if A is skew-hermitian

8 If A is hermitian and U is unitary, prove that $\bar{U}^TAU$ is hermitian

9 Under what conditions, if any, is it possible for every value of λ to be a characteristic value of the equation $(A - \lambda B)X = O$?

10 Show by an example that, if A and B are indefinite, the characteristic values of $(A - \lambda B)X = O$ need not be real, even though A and B are real and symmetric

11 Show that, if either A or B is nonsingular, then AB and BA have the same characteristic values Hence prove that there are no matrices A and B, with either A or B nonsingular, such that $AB - BA = I$ (These results hold even when both A and B are singular)

12 Show that the characteristic values of a real skew-symmetric matrix are either zero or pure imaginary

13 Prove that, if a (2,2) matrix has characteristic vectors which are orthogonal, it is symmetric

14 Prove that, if every characteristic value of a symmetric matrix is zero, the matrix is a null matrix Is this true if the matrix is not symmetric?

15 Show by an example that Corollary 1, Theorem 10, is false for symmetric matrices which are not real

11.3

The transformation of matrices

In previous sections we have already encountered the idea of the transformation of matrices For instance, in Sec 10 4 we defined equivalent matrices as matrices A and B connected by a relation of the form

$$B = QAP \qquad P,Q \text{ nonsingular}$$

Again, in Sec 11.1 we observed that, if the variables in a quadratic form X^TAX are subjected to a nonsingular linear transformation $X = PY$, then the quadratic form becomes

$$Y^TBY \qquad \text{where } B = P^TAP$$

In particular, we observed that, if a nonsingular matrix P can be found with the property that $B = P^TAP$ is a diagonal matrix, then, in terms of the new variables introduced by the substitution $X = PY$, the quadratic form X^TAX becomes just a sum of squares In this section we shall consider briefly the question of just when it is possible to transform an (n,n) matrix A into a diagonal matrix B by multiplying it on the right and on the left by suitable matrices P and Q.

Because an equivalence transformation is simply a composite of elementary transformations, it is clear that, for any square matrix A, many pairs of nonsingular matrices P and Q can be found such that QAP is diagonal In other words, we have the following result, which is little more than a restatement of Theorem 10, Sec 10 4, for square matrices

THEOREM 1

Any square matrix is equivalent to a diagonal matrix

COROLLARY 1

If a matrix A of rank r is equivalent to a diagonal matrix B, then B has exactly r nonzero diagonal elements

A square matrix cannot in general be transformed into a diagonal matrix by a transformation more restricted than an equivalence However, in many important special cases this is possible, as the following theorems show

THEOREM 2

A square matrix is congruent to a diagonal matrix if and only if it is symmetric

PROOF The proof that, if A is symmetric, then it is congruent to a diagonal matrix was essentially given in our discussion of the Lagrange reduction in Sec. 11 1 For, if A is symmetric, then it can be regarded as the matrix of a quadratic form, and the Lagrange reduction provides a linear transformation whose matrix P is a nonsingular matrix with the property that P^TAP is diagonal

On the other hand, if A is congruent to a diagonal matrix D, then there exists a nonsingular matrix P such that $A = P^TDP$ Furthermore, if this is the case, then the transpose of A is

$$A^T = (P^TDP)^T = P^TD^TP$$

However, $D^T = D$, since any diagonal matrix is obviously symmetric. Hence,

$$A^T = P^TD^TP = P^TDP = A$$

Therefore, A, being equal to its transpose, is symmetric, as asserted

From the nature of the Lagrange reduction it is evident that a symmetric matrix can be diagonalized by a congruence transformation in many ways, and among these there is always at least one which will simultaneously diagonalize a second given matrix, provided it is positive-definite More precisely, we have the following highly important theorem.

THEOREM 3

Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the (possibly repeated) characteristic values of the equation $(A - \lambda B)X = O$, where A and B are hermitian (or real symmetric) (n,n) matrices and B is positive-definite. Let $X_1, X_2, \ldots, X_n$ be n independent characteristic vectors corresponding to $\lambda_1, \lambda_2, \ldots, \lambda_n$, and let the X's be orthonormal with respect to B. Let M be the matrix whose columns are the characteristic vectors $X_1, X_2, \ldots, X_n$, and let D be the diagonal matrix whose diagonal elements are the characteristic values $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then $\bar{M}^T B M = I$, and $\bar{M}^T A M = D$.

PROOF Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the characteristic values of the equation $(A - \lambda B)X = O$. Whether or not there are repeated roots among the λ's, we know, from Corollary 2, Theorem 17, Sec. 11.2, that there exists a set of characteristic vectors $X_1, X_2, \ldots, X_n$ orthonormal with respect to B, that is, such that

$$(1) \qquad \bar{X}_i^T B X_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Now, writing the modal matrix M in partitioned form, for convenience, we have

$$(2) \qquad M = \begin{Vmatrix} X_1 & X_2 & \cdots & X_n \end{Vmatrix}$$

and

$$\bar{M}^T B M = \begin{Vmatrix} \bar{X}_1^T \\ \bar{X}_2^T \\ \cdot \\ \bar{X}_n^T \end{Vmatrix} \cdot \begin{Vmatrix} BX_1 & BX_2 & \cdots & BX_n \end{Vmatrix}$$

$$= \begin{Vmatrix} \bar{X}_1^T B X_1 & \bar{X}_1^T B X_2 & \cdots & \bar{X}_1^T B X_n \\ \bar{X}_2^T B X_1 & \bar{X}_2^T B X_2 & \cdots & \bar{X}_2^T B X_n \\ \cdot & \cdot & & \cdot \\ \bar{X}_n^T B X_1 & \bar{X}_n^T B X_2 & \cdots & \bar{X}_n^T B X_n \end{Vmatrix}$$

$$(3) \qquad = I \qquad \text{by (1)}$$

Also, by premultiplying Eq. (2) by A and then using the fact that for each i the X's are such that $AX_i = \lambda_i BX_i$, we have

$$\begin{aligned} AM &= \begin{Vmatrix} AX_1 & AX_2 & \cdots & AX_n \end{Vmatrix} \\ &= \begin{Vmatrix} \lambda_1 BX_1 & \lambda_2 BX_2 & \cdots & \lambda_n BX_n \end{Vmatrix} \\ &= B\begin{Vmatrix} \lambda_1 X_1 & \lambda_2 X_2 & \cdots & \lambda_n X_n \end{Vmatrix} \\ &= B\begin{Vmatrix} X_1 & X_2 & \cdots & X_n \end{Vmatrix} \cdot \begin{Vmatrix} \lambda_1 & & \bigcirc \\ & \lambda_2 & \\ & & \cdot \\ \bigcirc & & & \lambda_n \end{Vmatrix} \\ &= BMD \end{aligned}$$

Therefore, by (3),

$$\bar{M}^T A M = \bar{M}^T (BMD) = (\bar{M}^T B M) D = D$$

which is the second assertion of the theorem.

COROLLARY 1

If A and B are hermitian (or real symmetric) matrices, if B is positive-definite, and if M is a matrix whose columns are characteristic vectors of $(A - \lambda B)X = O$

orthonormal with respect to B, then the substitution $X = MY$ simultaneously reduces the hermitian (quadratic) forms $\bar{X}^TAX$ and $\bar{X}^TBX$ to $\bar{Y}^TDY$ and $\bar{Y}^TIY = \bar{Y}^TY$, respectively, where D is the diagonal matrix whose diagonal elements are the characteristic values which correspond respectively to the column vectors of M.

The conditions under which a square matrix can be diagonalized by a similarity transformation are contained in the next theorem

THEOREM 4

An (n,n) matrix is similar to a diagonal matrix if and only if it has n independent characteristic vectors.

PROOF Let A be an (n,n) matrix, and let us suppose first that A is similar to a diagonal matrix

$$D = \left\|\begin{matrix} d_{11} & & \bigcirc \\ & d_{22} & \\ \bigcirc & & d_{nn} \end{matrix}\right\|$$

that is, let us suppose that there exists a matrix S with the property that

$$(4) \qquad S^{-1}AS = D$$

If, for convenience, we write S in the partitioned form

$$S = \|S_1 \quad S_2 \quad \cdots \quad S_n\|$$

we have, by premultiplying Eq (4) by S,

$$AS = SD$$

or

$$\|AS_1 \quad AS_2 \quad \cdot \quad AS_n\| = \|S_1 \quad S_2 \quad \cdots \quad S_n\| \cdot \left\|\begin{matrix} d_{11} & & \bigcirc \\ & d_{22} & \\ \bigcirc & & d_{nn} \end{matrix}\right\|$$

$$= \|d_{11}S_1 \quad d_{22}S_2 \quad \cdot \quad d_{nn}S_n\|$$

Hence it follows that

$$AS_i = d_{ii}S_i = d_{ii}IS_i \qquad i = 1, 2, \ldots, n$$

which shows that $X = S_i$ is a characteristic vector corresponding to the characteristic value $\lambda_i = d_{ii}$ of the equation $(A - \lambda I)X = 0$ Thus the n columns of the transforming matrix S are characteristic vectors of the given matrix A Moreover, since the inverse of S exists, by hypothesis, it follows that $|S| \neq 0$ Hence, by Theorem 10, Sec 10 5, the n columns of S are linearly independent. Thus, the matrix A has n linearly independent characteristic vectors, and the necessity assertion of the theorem is verified

Suppose now that A has n linearly independent characteristic vectors X_1, X_2, $\ldots$, X_n corresponding to the (possibly repeated) characteristic values $\lambda_1, \lambda_2, \ldots, \lambda_n$ Then, by hypothesis,

$$AX_i = \lambda_i IX_i = \lambda_i X_i \qquad i = 1, 2, \ldots, n$$

Now, let S be the matrix whose columns are the characteristic vectors X_1, X_2,

. . . , X_n; i.e., let S be a modal matrix of A. Then, since the characteristic vectors are independent, by hypothesis, it follows that S^{-1} exists, and we can write

$$\begin{aligned} S^{-1}AS &= S^{-1}A\|X_1 \quad X_2 \quad \cdots \quad X_n\| \\ &= S^{-1}\ \|AX_1 \quad AX_2 \quad \cdot \quad AX_n\| \\ &= S^{-1}\ \|\lambda_1 X_1 \quad \lambda_2 X_2 \quad \cdots \quad \lambda_n X_n\| \\ &= S^{-1}\ \|X_1 \quad X_2 \quad\quad X_n\| \left\|\begin{matrix} \lambda_1 & & \bigcirc \\ & \lambda_2 & \\ & & \cdot \\ \bigcirc & & \lambda_n \end{matrix}\right\| \\ &= S^{-1}S \left\|\begin{matrix} \lambda_1 & & \bigcirc \\ & \lambda_2 & \\ & & \cdot \\ \bigcirc & & \lambda_n \end{matrix}\right\| \\ &= \left\|\begin{matrix} \lambda_1 & & \bigcirc \\ & \lambda_2 & \\ & & \cdot \\ \bigcirc & & \lambda_n \end{matrix}\right\| \end{aligned}$$

Hence, A is similar to a diagonal matrix; and the sufficiency assertion of the theorem is also verified.

Since every hermitian and every real symmetric matrix has n linearly independent characteristic vectors, it is clear that the last theorem contains the following important special result.

COROLLARY 1

Every hermitian and every real symmetric matrix is similar to a diagonal matrix.

Using the Schmidt process, it is clear that if a matrix A has n independent characteristic vectors, it has, in fact, a set of n orthonormal characteristic vectors. Moreover, as we saw in Exercise 12, Sec. 10.3, a matrix whose columns are orthonormal is an orthogonal matrix. Hence, taking the matrix S in Theorem 4 to be a matrix whose columns are orthonormal characteristic vectors of A, we have the following results.

COROLLARY 2

Every real symmetric matrix is orthogonally similar to a diagonal matrix.

COROLLARY 3

If a matrix is orthogonally similar to a diagonal matrix, it is symmetric.

By essentially the same argument, the following companion results for hermitian matrices can be established.

COROLLARY 4

Every hermitian matrix is unitarily similar to a diagonal matrix.

COROLLARY 5

If a matrix is unitarily similar to a diagonal matrix, it is hermitian.

Although not every matrix is similar to a diagonal matrix, every matrix is similar to a triangular matrix Specifically, in more advanced texts* the following results are established:

THEOREM 5

Every square matrix is unitarily similar to a triangular matrix

THEOREM 6

Let the characteristic values of an (n,n) matrix A be $\lambda_1, \lambda_2, \ldots, \lambda_k$, let the multiplicity of λ_i be m_i; let r_i be the number of linearly independent characteristic vectors of A corresponding to λ_i, and let D_i be the (m_i,m_i) upper triangular matrix in which the diagonal elements are all λ_i, the first $m_i - r_i$ elements on the diagonal above the principal diagonal are each 1, and all other elements are 0. Then the given matrix is similar to the matrix

$$D = \left\| \begin{matrix} D_1 & & \bigcirc \\ & D_2 & \\ \bigcirc & & D_k \end{matrix} \right\|$$

The standard, or canonical, form described in the last theorem is known as the **Jordan canonical form.**†

Many of the theorems of the last two sections find their most immediate physical application in the analysis of vibrating systems, either mechanical or electrical In particular, the orthogonality of the characteristic vectors of a matric equation make it possible to impose initial conditions of displacement and velocity on a system with a finite number of degrees of freedom in a way that resembles closely the corresponding procedure for boundary value problems involving continuous systems (Secs 8 4 and 8 5) The following example illustrates these ideas

EXAMPLE 1

The three masses shown in Fig 11 1a are initially displaced so that

$$(x_1)_0 = 2 \qquad (x_2)_0 = -1 \qquad (x_3)_0 = 1$$

From these positions they begin to move with initial velocities

$$(v_1)_0 = 0 \qquad (v_2)_0 = 2 \qquad (v_3)_0 = 0$$

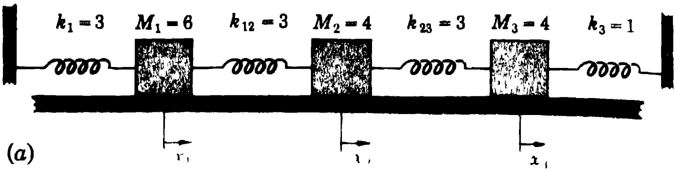
Assuming that there is no friction in the system, determine the subsequent motion of each mass

Since friction is assumed to be negligible, the only forces acting are those transmitted to the masses by the springs directly attached to them Now, when the instantaneous displacements of the masses are x_1, x_2, and x_3, the lengths of the springs have changed from their unstretched,

---

* See, for instance, L Mirsky, "Linear Algebra," p. 307, Oxford Book Company, Inc , New York, 1955.

† Named for the French mathematician Camille Jordan (1838–1922).



Net changes in spring lengths (x_1, x_3 shown positive, x_2 shown negative)

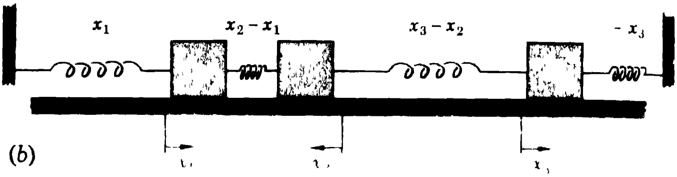


FIGURE 11.1
A three-mass system in equilibrium and in a displaced position.

equilibrium lengths by the respective amounts (Fig. 11.1b)

$$x_1 \qquad x_2 - x_1 \qquad x_3 - x_2 \qquad -x_3$$

Hence, the forces instantaneously exerted by the springs are, respectively,

$$3x_1 \qquad 3(x_2 - x_1) \qquad 3(x_3 - x_2) \qquad -x_3$$

where plus signs indicate that the springs are in tension, minus signs that the springs are in compression. Therefore, applying Newton's law to each of the masses in turn, we obtain the three differential equations

$$6\frac{d^2x_1}{dt^2} = -3x_1 + 3(x_2 - x_1)$$

$$4\frac{d^2x_2}{dt^2} = -3(x_2 - x_1) + 3(x_3 - x_2)$$

$$4\frac{d^2x_3}{dt^2} = -3(x_3 - x_2) - x_3$$

or

$$\begin{aligned} (6D^2 + 6)x_1 - 3x_2 \qquad\qquad &= 0 \\ -3x_1 + (4D^2 + 6)x_2 - 3x_3 &= 0 \\ -3x_2 + (4D^2 + 4)x_3 &= 0 \end{aligned} \tag{5}$$

or, in matrix notation, simply

$$P(D)X = O$$

where
$$P(D) = \begin{Vmatrix} 6D^2 + 6 & -3 & 0 \\ -3 & 4D^2 + 6 & -3 \\ 0 & -3 & 4D^2 + 4 \end{Vmatrix} \qquad \text{and} \qquad X = \begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix}$$

Since there is no dissipation of energy through friction, it is clear that each mass must vibrate around its equilibrium position with constant amplitude. Hence, as a solution we assume

$$X = \begin{Vmatrix} a_1 \\ a_2 \\ a_3 \end{Vmatrix} \cos \omega t$$

that is,

$$x_1 = a_1 \cos \omega t \qquad x_2 = a_2 \cos \omega t \qquad x_3 = a \cos \omega t$$

where ω is an unknown frequency and a_1, a_2, and a_3 are the unknown amplitudes through which the masses oscillate. Substituting these into the differential equations in (5) and dividing out the common factor $\cos \omega t$, we obtain the three algebraic equations

$$
(6)\qquad
\begin{aligned}
(-6\omega^2 + 6)a_1 - {} & 3a_2 & & = 0\\
-3a_1 + {} & (-4\omega^2 + 6)a_2 - {} & 3a_3 & = 0\\
& -3a_2 + {} & (-4\omega^2 + 4)a_3 & = 0
\end{aligned}
$$

from which to determine a_1, a_2, and a_3. This system will have a nontrivial solution if and only if the determinant of its coefficients is equal to zero. Hence, we must have

$$
\begin{vmatrix} (-6\omega^2+6) & -3 & 0\\ -3 & (-4\omega^2+6) & -3\\ 0 & -3 & (-4\omega^2+4)\end{vmatrix} = -6(4\omega^2-1)(\omega^2-1)(4\omega^2-9) = 0
$$

Thus the system (6) has a nontrivial solution for $\omega^2 = \frac{1}{4}, 1, \frac{9}{4}$ and for no other values of ω^2. The natural frequencies of the physical system are, therefore,

$$\omega = \tfrac{1}{2}, 1, \tfrac{3}{2}$$

Now, according to Theorem 8, Sec. 10.5, the values of a_1, a_2, and a_3 which satisfy (6) when the determinant of its coefficients is equal to zero can be read from any (2,3) matrix of rank 2 contained in the coefficient matrix. Hence, using the matrix of the coefficients of the last two equations in the set, we have, in the three nontrivial cases,

$$
\omega = \tfrac{1}{2} \qquad \left\|\begin{matrix} -3 & 5 & -3\\ 0 & -3 & 3\end{matrix}\right\| \qquad \frac{a_1}{\begin{vmatrix} 5 & -3\\ -3 & 3\end{vmatrix}} = \frac{a_2}{-\begin{vmatrix} -3 & -3\\ 0 & 3\end{vmatrix}} = \frac{a_3}{\begin{vmatrix} -3 & 5\\ 0 & -3\end{vmatrix}}
$$

$$a_1 = 2 \qquad a_2 = 3 \qquad a_3 = 3$$

$$
\omega = 1 \qquad \left\|\begin{matrix} -3 & 2 & -3\\ 0 & -3 & 0\end{matrix}\right\| \qquad \frac{a_1}{\begin{vmatrix} 2 & -3\\ -3 & 0\end{vmatrix}} = \frac{a_2}{-\begin{vmatrix} -3 & -3\\ 0 & 0\end{vmatrix}} = \frac{a_3}{\begin{vmatrix} -3 & 2\\ 0 & -3\end{vmatrix}}
$$

$$a_1 = 1 \qquad a_2 = 0 \qquad a_3 = -1$$

$$
\omega = \tfrac{3}{2} \qquad \left\|\begin{matrix} -3 & -3 & -3\\ 0 & -3 & -5\end{matrix}\right\| \qquad \frac{a_1}{\begin{vmatrix} -3 & -3\\ -3 & -5\end{vmatrix}} = \frac{a_2}{-\begin{vmatrix} -3 & -3\\ 0 & -5\end{vmatrix}} = \frac{a_3}{\begin{vmatrix} -3 & -3\\ 0 & -3\end{vmatrix}}
$$

$$a_1 = 2 \qquad a_2 = -5 \qquad a_3 = 3$$

Thus we have found three particular solution vectors for the system (5), namely,

$$
X_1 = \left\|\begin{matrix}2\\3\\3\end{matrix}\right\| \cos\frac{t}{2} \qquad X_2 = \left\|\begin{matrix}1\\0\\-1\end{matrix}\right\| \cos t \qquad X_3 = \left\|\begin{matrix}2\\-5\\3\end{matrix}\right\| \cos\frac{3}{2}t
$$

Clearly, if we had begun with the assumptions

$$x_1 = a_1 \sin \omega t \qquad x_2 = a_2 \sin \omega t \qquad x_3 = a_3 \sin \omega t$$

we would also have obtained the algebraic equations (6) and, hence, the same three values of ω and the same solution vectors. Therefore, we have three more particular solutions:

$$
X_4 = \left\|\begin{matrix}2\\3\\3\end{matrix}\right\| \sin\frac{t}{2} \qquad X_5 = \left\|\begin{matrix}1\\0\\-1\end{matrix}\right\| \sin t \qquad X_6 = \left\|\begin{matrix}2\\-5\\3\end{matrix}\right\| \sin\frac{3}{2}t
$$

and, finally, the complete solution

$$(7)\qquad X = c_1X_1 + c_2X_2 + c_3X_3 + c_4X_4 + c_5X_5 + c_6X_6$$

where the c's are arbitrary scalar coefficients.

To determine the values of the c's we must, of course, use the given initial conditions. The most convenient way to do this is to write the system (6) in the form

$$(8)\qquad (V - \omega^2 T)A = O$$

where
$$V = \begin{Vmatrix} 6 & -3 & 0 \\ -3 & 6 & -3 \\ 0 & -3 & 4 \end{Vmatrix} \qquad T = \begin{Vmatrix} 6 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{Vmatrix} \qquad A = \begin{Vmatrix} a_1 \\ a_2 \\ a_3 \end{Vmatrix}$$

and then recall from Sec 11.2 that the solution vectors of (8), namely,

$$A_1 = \begin{Vmatrix} 2 \\ 3 \\ 3 \end{Vmatrix} \qquad A_2 = \begin{Vmatrix} 1 \\ 0 \\ -1 \end{Vmatrix} \qquad A_3 = \begin{Vmatrix} 2 \\ -5 \\ 3 \end{Vmatrix}$$

satisfy the orthogonality condition

$$A_i^T T A_j = 0 \qquad i \neq j \tag{9}$$

To take advantage of this property, we first set $t = 0$ in (7) and substitute the initial displacement vector for $X(0)$, getting

$$\begin{Vmatrix} 2 \\ -1 \\ 1 \end{Vmatrix} = c_1 \begin{Vmatrix} 2 \\ 3 \\ 3 \end{Vmatrix} + c_2 \begin{Vmatrix} 1 \\ 0 \\ -1 \end{Vmatrix} + c_3 \begin{Vmatrix} 2 \\ -5 \\ 3 \end{Vmatrix} \tag{10}$$

Then, if we multiply this equation through on the left by

$$A_1^T T = \begin{Vmatrix} 2 & 3 & 3 \end{Vmatrix} \begin{Vmatrix} 6 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{Vmatrix} = \begin{Vmatrix} 12 & 12 & 12 \end{Vmatrix}$$

the second and third terms on the right vanish because of the orthogonality property (9), and we have simply

$$\begin{Vmatrix} 12 & 12 & 12 \end{Vmatrix} \begin{Vmatrix} 2 \\ -1 \\ 1 \end{Vmatrix} = c_1 \begin{Vmatrix} 12 & 12 & 12 \end{Vmatrix} \begin{Vmatrix} 2 \\ 3 \\ 3 \end{Vmatrix} \qquad \text{or} \qquad c_1 = \tfrac{1}{4}$$

Similarly, multiplying (10) on the left by

$$A_2^T T = \begin{Vmatrix} 6 & 0 & -4 \end{Vmatrix} \quad \text{and by} \quad A_3^T T = \begin{Vmatrix} 12 & -20 & 12 \end{Vmatrix}$$

in turn, we find

$$c_2 = \tfrac{4}{5} \qquad c_3 = \tfrac{7}{20}$$

To find c_4, c_5, and c_6 we first differentiate Eq (7), getting

$$\begin{aligned} \frac{dX}{dt} = &-\frac{1}{2}c_1 \begin{Vmatrix} 2 \\ 3 \\ 3 \end{Vmatrix} \sin\frac{t}{2} - c_2 \begin{Vmatrix} 1 \\ 0 \\ -1 \end{Vmatrix} \sin t - \frac{3}{2}c_3 \begin{Vmatrix} 2 \\ -5 \\ 3 \end{Vmatrix} \sin\frac{3}{2}t \\ &+ \frac{1}{2}c_4 \begin{Vmatrix} 2 \\ 3 \\ 3 \end{Vmatrix} \cos\frac{t}{2} + c_5 \begin{Vmatrix} 1 \\ 0 \\ -1 \end{Vmatrix} \cos t + \frac{3}{2}c_6 \begin{Vmatrix} 2 \\ -5 \\ 3 \end{Vmatrix} \cos\frac{3}{2}t \end{aligned}$$

Then, setting $t = 0$ and replacing $\left.\dfrac{dX}{dt}\right|_{t=0}$ by the given initial velocity vector $\begin{Vmatrix} 1 \\ 2 \\ 0 \end{Vmatrix}$, we have

$$\begin{Vmatrix} 1 \\ 2 \\ 0 \end{Vmatrix} = \tfrac{1}{2}c_4 \begin{Vmatrix} 2 \\ 3 \\ 3 \end{Vmatrix} + c_5 \begin{Vmatrix} 1 \\ 0 \\ -1 \end{Vmatrix} + \tfrac{3}{2}c_6 \begin{Vmatrix} 2 \\ -5 \\ 3 \end{Vmatrix} \tag{11}$$

Finally, multiplying this equation on the left by

$$A_1^T T = \begin{Vmatrix} 12 & 12 & 12 \end{Vmatrix} \qquad A_2^T T = \begin{Vmatrix} 6 & 0 & -4 \end{Vmatrix} \qquad \text{and} \qquad A_3^T T = \begin{Vmatrix} 12 & -20 & 12 \end{Vmatrix}$$

in turn, we find

$$c_4 = \tfrac{3}{4} \qquad c_5 = \tfrac{3}{5} \qquad \text{and} \qquad c_6 = -\tfrac{7}{40}$$

With the c's determined, the solution is now complete, and we have

$$X = \tfrac{1}{4}X_1 + \tfrac{4}{5}X_2 + \tfrac{7}{20}X_3 + \tfrac{3}{4}X_4 + \tfrac{3}{5}X_5 - \tfrac{7}{40}X_6$$

or, explicitly,

$$\begin{aligned}
x_1 &= \frac{1}{2}\cos\frac{t}{2} + \frac{4}{5}\cos t + \frac{7}{10}\cos\frac{3}{2}t + \frac{3}{2}\sin\frac{t}{2} + \frac{3}{5}\sin t - \frac{7}{20}\sin\frac{3}{2}t \\
x_2 &= \frac{3}{4}\cos\frac{t}{2} \qquad\qquad - \frac{7}{4}\cos\frac{3}{2}t + \frac{9}{4}\sin\frac{t}{2} \qquad\qquad + \frac{7}{8}\sin\frac{3}{2}t \\
x_3 &= \frac{3}{4}\cos\frac{t}{2} - \frac{4}{5}\cos t + \frac{21}{20}\cos\frac{3}{2}t + \frac{9}{4}\sin\frac{t}{2} - \frac{3}{5}\sin t - \frac{21}{40}\sin\frac{3}{2}t
\end{aligned}$$

We have already identified the three values $\omega = \frac{1}{2}, 1, \frac{3}{2}$ as the natural frequencies of the system, i.e., the only frequencies at which free vibrations of the system are possible, and we have illustrated how the motion produced by an arbitrary set of initial conditions involves simultaneously vibrations at each of the natural frequencies. The vectors A_1, A_2, and A_3, associated, respectively, with the frequencies $\omega = \frac{1}{2}$, $\omega = 1$, and $\omega = \frac{3}{2}$, are called the **normal modes** of the system. Each describes the relative amplitudes with which the three masses would vibrate if the system were set in motion in such a way that it vibrated only at the corresponding natural frequency. The *absolute* amplitudes depend upon the c's, of course, and so are determined by the initial conditions, but at each natural frequency the *ratios* of the amplitudes with which the masses oscillate are always the same, regardless of their actual numerical values. Figure 11.2 illustrates this behavior for one full cycle of the motion at each of the three natural frequencies.

To conclude our discussion, let us now apply to this problem the results of Theorem 3, Sec. 11.3. To do this, we return to the matric equation (8), namely, $(V - \omega^2 T)A = O$, and observe that V and T are both symmetric and that T is positive-definite. Hence, the hypotheses of Corollary 1, Theorem 3, Sec. 11.3, are fulfilled, therefore, if A_1^*, A_2^*, A_3^* are solution vectors of Eq. (8) orthonormal with respect to T and if M is the matrix $\|A_1^* \quad A_2^* \quad A_3^*\|$, the substitution

$$X = MY$$

will simultaneously reduce the quadratic forms X^TVX and X^TTX to the respective diagonal forms Y^TDY and Y^TY, where D is the diagonal matrix of characteristic values

$$\left\|\begin{matrix} \frac{1}{4} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{9}{4} \end{matrix}\right\|$$

To verify this, we note first that, when the solution vectors

$$A_1 = \left\|\begin{matrix} 2 \\ 3 \\ 3 \end{matrix}\right\| \qquad A_2 = \left\|\begin{matrix} 1 \\ 0 \\ -1 \end{matrix}\right\| \qquad A_3 = \left\|\begin{matrix} 2 \\ -5 \\ 3 \end{matrix}\right\|$$

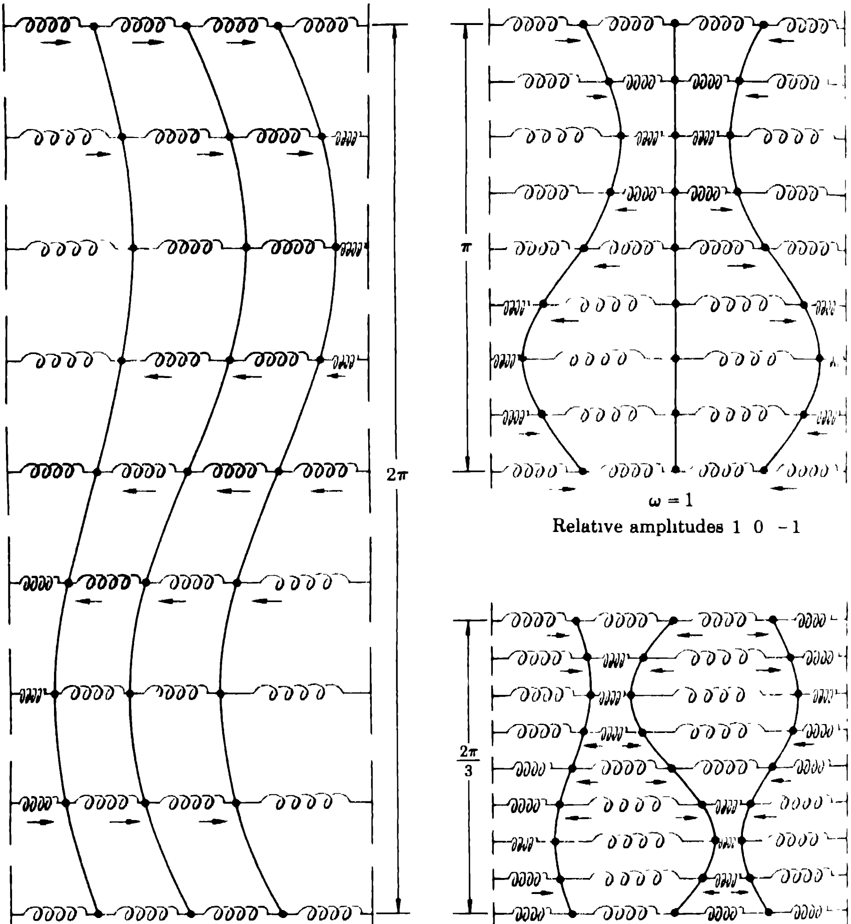
are normalized with respect to T, we obtain

$$A_1^* = \frac{A_1}{\sqrt{A_1{}^TTA_1}} = \frac{1}{\sqrt{96}}\left\|\begin{matrix} 2 \\ 3 \\ 3 \end{matrix}\right\| \qquad A_2^* = \frac{A_2}{\sqrt{A_2{}^TTA_2}} = \frac{1}{\sqrt{10}}\left\|\begin{matrix} 1 \\ 0 \\ -1 \end{matrix}\right\|$$

$$A_3^* = \frac{A_3}{\sqrt{A_3{}^TTA_3}} = \frac{1}{\sqrt{160}}\left\|\begin{matrix} 2 \\ -5 \\ 3 \end{matrix}\right\|$$

Hence, the required substitution $X = MY$ is

$$\left\|\begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix}\right\| = \left\|\begin{matrix} \frac{2}{\sqrt{96}} & \frac{1}{\sqrt{10}} & \frac{2}{\sqrt{160}} \\ \frac{3}{\sqrt{96}} & 0 & \frac{-5}{\sqrt{160}} \\ \frac{3}{\sqrt{96}} & \frac{-1}{\sqrt{10}} & \frac{3}{\sqrt{160}} \end{matrix}\right\| \left\|\begin{matrix} y_1 \\ y_2 \\ y_3 \end{matrix}\right\|$$



$\omega = 1$

Relative amplitudes 1 0 −1

$\omega = \frac{1}{2}$

Relative amplitudes 2 3 3

$\omega = \frac{3}{2}$

Relative amplitudes 2 −5 3

FIGURE 11.2

The normal modes of the system shown in Fig. 11 1

or

$$
\begin{aligned}
x_1 &= \frac{2y_1}{\sqrt{96}} + \frac{y_2}{\sqrt{10}} + \frac{2y_3}{\sqrt{160}} \\
x_2 &= \frac{3y_1}{\sqrt{96}} \qquad\qquad - \frac{5y_3}{\sqrt{160}} \\
x_3 &= \frac{3y_1}{\sqrt{96}} - \frac{y_2}{\sqrt{10}} + \frac{3y_3}{\sqrt{160}}
\end{aligned}
\tag{12}
$$

Finally, introducing these expressions into the quadratic forms

$$X^TVX = 6x_1^2 - 6x_1x_2 + 6x_2^2 - 6x_2x_3 + 4x_3^2 \quad \text{and} \quad X^TTX = 6x_1^2 + 4x_2^2 + 4x_3^2$$

we obtain, respectively,

$$\begin{aligned}
X^TVX = {} & 6\left(\frac{2y_1}{\sqrt{96}} + \frac{y_2}{\sqrt{10}} + \frac{2y_3}{\sqrt{160}}\right)^2 \\
& - 6\left(\frac{2y_1}{\sqrt{96}} + \frac{y_2}{\sqrt{10}} + \frac{2y_3}{\sqrt{160}}\right)\left(\frac{3y_1}{\sqrt{96}} - \frac{5y_3}{\sqrt{160}}\right) \\
& + 6\left(\frac{3y_1}{\sqrt{96}} - \frac{5y_3}{\sqrt{160}}\right)^2 \\
& - 6\left(\frac{3y_1}{\sqrt{96}} - \frac{5y_3}{\sqrt{160}}\right)\left(\frac{3y_1}{\sqrt{96}} - \frac{y_2}{\sqrt{10}} + \frac{3y_3}{\sqrt{160}}\right) \\
& \qquad\qquad + 1\left(\frac{3y_1}{\sqrt{96}} - \frac{y_2}{\sqrt{10}} + \frac{3y_3}{\sqrt{160}}\right)^2 \\
= {} & \frac{1}{4}y_1^2 + y_2^2 + \frac{9}{4}y_3^2
\end{aligned}$$

and

$$\begin{aligned}
X^TTX = {} & 6\left(\frac{2y_1}{\sqrt{96}} + \frac{y_2}{\sqrt{10}} + \frac{2y_3}{\sqrt{160}}\right)^2 + 4\left(\frac{3y_1}{\sqrt{96}} - \frac{5y_3}{\sqrt{160}}\right)^2 \\
& \qquad\qquad + 4\left(\frac{3y_1}{\sqrt{96}} - \frac{y_2}{\sqrt{10}} + \frac{3y_3}{\sqrt{160}}\right)^2 \\
= {} & y_1^2 + y_2^2 + y_3^2
\end{aligned}$$

In the present problem it is easy to identify the two quadratic forms X^TVX and X^TTX. In fact, since the energy stored in a spring stretched a distance s is $ks^2/2$, it follows that the instantaneous potential energy in our system is

$$\begin{aligned}
\tfrac{1}{2}[3x_1^2 + 3(x_2 - x_1)^2 + 3(x_3 - x_2)^2 + x_1^2] &= \tfrac{1}{2}[6x_1^2 - 6x_1x_2 + 6x_2^2 - 6x_2x_3 + 4x_3^2] \\
&= \tfrac{1}{2}X^TVX
\end{aligned}$$

Hence, X^TVX is equal to twice the instantaneous potential energy of the system.

Also, the kinetic energy of a mass moving with velocity v is $mv^2/2$. Hence, the instantaneous kinetic energy of our system is

$$\tfrac{1}{2}(6\dot{x}_1^2 + 4\dot{x}_2^2 + 4\dot{x}_3^2) = \tfrac{1}{2}\dot{X}^TT\dot{X}$$

From this it follows that, when the system is vibrating at any one of its natural frequencies ω_i, its maximum kinetic energy is

$$\frac{\omega_i^2}{2}X^TTX$$

The new coordinates y_1, y_2, y_3, defined by (12) and in terms of which the two energy expressions appear as sums of squares, are known as the **normal coordinates** of the system.

EXERCISES

1 For each of the following matrices A, find two pairs of nonsingular matrices (P,Q) such that PAQ is a diagonal matrix.

a $\left\|\begin{matrix} 1 & 2 \\ 3 & 4 \end{matrix}\right\|$ **b** $\left\|\begin{matrix} 1 & -1 \\ 0 & 3 \end{matrix}\right\|$ **c** $\left\|\begin{matrix} 1 & -1 & 1 \\ 2 & 1 & 2 \\ 0 & 1 & 3 \end{matrix}\right\|$ **d** $\left\|\begin{matrix} 1 & 0 & 3 \\ 1 & -1 & 1 \\ -1 & 3 & 3 \end{matrix}\right\|$

2 For each of the following matrices A, find two nonsingular matrices P such that P^TAP is a diagonal matrix

a $\begin{Vmatrix} 1 & 2 \\ 2 & 3 \end{Vmatrix}$ **b** $\begin{Vmatrix} 1 & -1 \\ -1 & 0 \end{Vmatrix}$ **c** $\begin{Vmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 3 \end{Vmatrix}$ **d** $\begin{Vmatrix} 1 & 2 & 0 \\ 2 & 5 & 2 \\ 0 & 2 & 4 \end{Vmatrix}$

3 For each of the following pairs of matrices (A,B), find a congruence transformation which will simultaneously reduce A and B to diagonal form, and carry out the diagonalization.

a $\begin{Vmatrix} 3 & -2 \\ -2 & 4 \end{Vmatrix}$ $\begin{Vmatrix} 1 & 0 \\ 0 & 2 \end{Vmatrix}$ **b** $\begin{Vmatrix} 6 & 2 \\ 2 & 2 \end{Vmatrix}$ $\begin{Vmatrix} 2 & 0 \\ 0 & 1 \end{Vmatrix}$

c $\begin{Vmatrix} 4 & 3 \\ 3 & 6 \end{Vmatrix}$ $\begin{Vmatrix} 1 & 0 \\ 0 & 3 \end{Vmatrix}$ **d** $\begin{Vmatrix} 2 & 2 \\ 2 & 3 \end{Vmatrix}$ $\begin{Vmatrix} 1 & 1 \\ 1 & 2 \end{Vmatrix}$

e $\begin{Vmatrix} 8 & -2 & 0 \\ -2 & 3 & -1 \\ 0 & -1 & 2 \end{Vmatrix}$ $\begin{Vmatrix} 8 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{Vmatrix}$ **f** $\begin{Vmatrix} 3 & -1 & 0 \\ -1 & 1 & -1 \\ 0 & -1 & 5 \end{Vmatrix}$ $\begin{Vmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{Vmatrix}$

g $\begin{Vmatrix} 6 & 0 & 2 \\ 0 & 6 & -4 \\ 2 & -4 & 6 \end{Vmatrix}$ $\begin{Vmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{Vmatrix}$ **h** $\begin{Vmatrix} 7 & -1 & 0 \\ -1 & 11 & -4 \\ 0 & -4 & 10 \end{Vmatrix}$ $\begin{Vmatrix} 6 & -2 & -1 \\ -2 & 10 & -5 \\ -1 & -5 & 9 \end{Vmatrix}$

4 **a** If A and B are hermitian (or real symmetric) matrices, show that there may exist congruence transformations which will simultaneously diagonalize A and B even though B is not definite.

b Find a congruence transformation which will simultaneously diagonalize $\begin{Vmatrix} -2 & 1 \\ 1 & 1 \end{Vmatrix}$ and $\begin{Vmatrix} 1 & 0 \\ 0 & -2 \end{Vmatrix}$.

c Find a congruence transformation which will simultaneously diagonalize $\begin{Vmatrix} -3 & 3 \\ 3 & 0 \end{Vmatrix}$ and $\begin{Vmatrix} -7 & 5 \\ 5 & -1 \end{Vmatrix}$

5 Find similarity transformations which will reduce each of the following matrices to diagonal form

a $\begin{Vmatrix} -3 & 2 \\ -10 & 6 \end{Vmatrix}$ **b** $\begin{Vmatrix} 0 & -2 \\ -2 & 0 \end{Vmatrix}$ **c** $\begin{Vmatrix} 2 & 1 \\ 2 & 1 \end{Vmatrix}$

d $\begin{Vmatrix} 5 & -2 & -1 \\ -1 & 4 & -1 \\ 1 & -2 & 3 \end{Vmatrix}$ **e** $\begin{Vmatrix} 2 & -3 & 3 \\ 0 & 3 & -1 \\ 0 & -1 & 3 \end{Vmatrix}$ **f** $\begin{Vmatrix} 3 & -2 & -2 \\ -1 & 2 & 0 \\ 1 & -1 & 1 \end{Vmatrix}$

6 Work Example 1 with $X_0 = \begin{Vmatrix} 1 \\ 2 \\ 2 \end{Vmatrix}$ and $\dot{X}_0 = \begin{Vmatrix} 1 \\ -1 \\ 3 \end{Vmatrix}$

7 The system shown in Fig 11 3 begins to move with initial displacement $X_0 = \begin{Vmatrix} 1 \\ 2 \end{Vmatrix}$ and

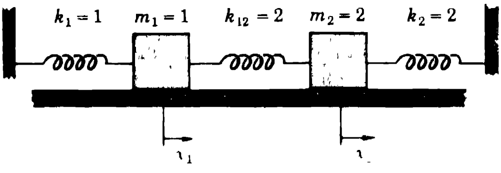
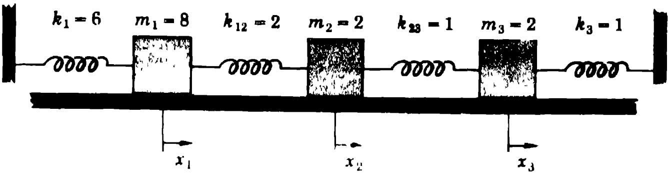


FIGURE 11.3

initial velocity $\dot{X}_0 = \begin{Vmatrix} 2 \\ -1 \end{Vmatrix}$ Assuming that there is no friction in the system, determine its subsequent motion

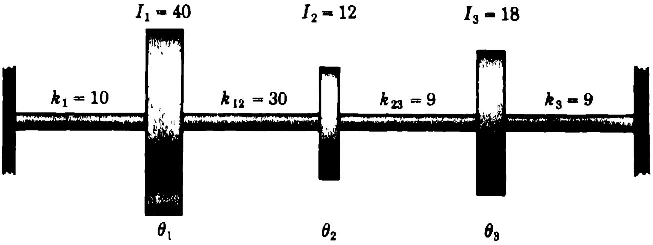
8 The system shown in Fig. 11.4 begins to move with initial displacement $X_0 = \begin{Vmatrix} 1 \\ 1 \\ -1 \end{Vmatrix}$ and initial velocity $\dot{X}_0 = \begin{Vmatrix} 1 \\ 0 \\ 2 \end{Vmatrix}$. Assuming that there is no friction in the system, determine its subsequent motion.

FIGURE 11.4



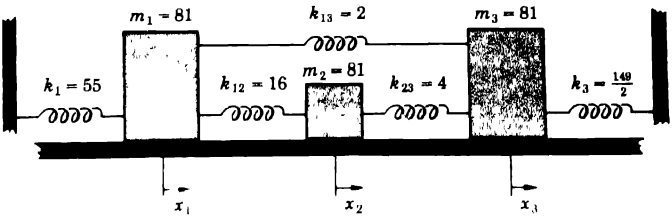
9 The system shown in Fig. 11.5 begins to move with initial displacement $\theta = \begin{Vmatrix} 1 \\ 0 \\ 2 \end{Vmatrix}$ and initial velocity $\dot{\theta} = \begin{Vmatrix} 2 \\ 1 \\ 1 \end{Vmatrix}$. Assuming that there is no friction in the system, determine its subsequent motion.

FIGURE 11.5



10 Find the natural frequencies and normal modes of the system shown in Fig. 11.6.

FIGURE 11.6



11.4

Functions of a square matrix

In Sec. 10.2, after we had defined matric multiplication, we were able to define positive integral powers of a square matrix A and to verify that, for arbitrary positive integers r and s,

(1) $$A^rA^s = A^sA^r = A^{r+s}$$

Moreover, we verified in Sec. 10 3 that, if A is a nonsingular matrix, it has an inverse A^{-1} such that $AA^{-1} = A^{-1}A = I$, and we defined negative integral powers of A by the relation

$$A^{-n} = (A^{-1})^n$$

Thus, after we introduced the definition $A^0 = I$, it became clear that, for any nonsingular matrix A, Eq. (1) holds for all integral values of r and s It is now natural to define polynomial functions of a square matrix and, if possible, rational fractional functions

DEFINITION 1

A polynomial function of a square matrix A is a finite linear combination of nonnegative integral powers of A,

$$p(A) = a_0A^n + a_1A^{n-1} + \qquad + a_{n-1}A + a_nI$$

EXAMPLE 1

If

$$A = \left\| \begin{matrix} 1 & 2 \\ 3 & -4 \end{matrix} \right\| \qquad \text{and} \qquad p(x) = x^2 + 5x + 4$$

then

$$p(A) = A^2 + 5A + 4I = \left\| \begin{matrix} 7 & -6 \\ -9 & 22 \end{matrix} \right\| + 5\left\| \begin{matrix} 1 & 2 \\ 3 & -4 \end{matrix} \right\| + 4\left\| \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \right\| = \left\| \begin{matrix} 16 & 4 \\ 6 & 6 \end{matrix} \right\|$$

It is interesting to note that $p(A)$ can also be evaluated by using the factored forms of $p(x)$, namely,

$$p(x) = (x + 4)(x + 1) = (x + 1)(x + 4)$$

$$p(A) = (A + 4I)(A + I) = \left(\left\| \begin{matrix} 1 & 2 \\ 3 & -4 \end{matrix} \right\| + 4\left\| \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \right\| \right) \left(\left\| \begin{matrix} 1 & 2 \\ 3 & -4 \end{matrix} \right\| + \left\| \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \right\| \right)$$

$$= \left\| \begin{matrix} 5 & 2 \\ 3 & 0 \end{matrix} \right\| \left\| \begin{matrix} 2 & 2 \\ 3 & -3 \end{matrix} \right\| = \left\| \begin{matrix} 16 & 4 \\ 6 & 6 \end{matrix} \right\|$$

$$p(A) = (A + I)(A + 4I) = \left(\left\| \begin{matrix} 1 & 2 \\ 3 & -4 \end{matrix} \right\| + \left\| \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \right\| \right) \left(\left\| \begin{matrix} 1 & 2 \\ 3 & -4 \end{matrix} \right\| + 4\left\| \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \right\| \right)$$

$$= \left\| \begin{matrix} 2 & 2 \\ 3 & -3 \end{matrix} \right\| \left\| \begin{matrix} 5 & 2 \\ 3 & 0 \end{matrix} \right\| = \left\| \begin{matrix} 16 & 4 \\ 6 & 6 \end{matrix} \right\|$$

In this example it is of course not clear, especially in view of the noncommutative character of matric multiplication, whether the fact that $p(A)$ can be computed equally well from

$$A^2 + 5A + 4I \qquad (A + 4I)(A + I) \qquad \text{and} \qquad (A + I)(A + 4I)$$

is a result of some special property of A and $p(x)$ or is illustrative of some general principle Actually the latter is the case, in fact, any identical relation involving sums and products of scalar polynomials is valid for the corresponding matric polynomials, as the following important, but almost obvious, theorem assures us

THEOREM 1

Any polynomial identity between scalar polynomials implies a corresponding identity for matric polynomials.

PROOF Clearly, any polynomial relation between scalar polynomials can be constructed using only the operations of addition and multiplication. For instance,

$$[f_1(x)f_2(x) + f_3(x)]f_4(x) = f_5(x)$$

is completely equivalent to the chain of relations

$$\phi(x)f_4(x) = f_5(x) \qquad \phi(x) = \psi(x) + f_3(x) \qquad \psi(x) = f_1(x)f_2(x)$$

Hence, to prove the theorem it is sufficient to show that for any polynomials f, g, s, p and any square matrix A,

a If $f(x) + g(x) = s(x)$, then $f(A) + g(A) = s(A)$

b If $f(x)g(x) = p(x)$, then $f(A)g(A) = p(A)$

To prove the first of these, let

$$f(x) = \sum_{i=0}^{m} a_i x^i \qquad g(x) = \sum_{i=0}^{n} b_i x^i \qquad s(x) = \sum_{i=0}^{t} c_i x^i$$

where $t = \max(m,n)$, $c_i = a_i + b_i$, and the coefficients of any powers of x which are not present are understood to be zero. Then

$$\begin{aligned} f(A) + g(A) &= \sum_{i=0}^{m} a_i A^i + \sum_{i=0}^{n} b_i A^i \\ &= \sum_{i=0}^{t} (a_i + b_i)A^i = \sum_{i=0}^{t} c_i A^i = s(A) \qquad \text{as asserted} \end{aligned}$$

To prove part **b**, let

$$f(x) = \sum_{i=0}^{m} a_i x^i \qquad g(x) = \sum_{j=0}^{n} b_j x^j \qquad p(x) = \sum_{k=0}^{t} c_k x^k$$

where $t = m + n$ and $c_k = \sum_{i,j} a_i b_j$, the summation extending over all values of i and j such that $i + j = k$ and, of course, $0 \leqq i \leqq m$, $0 \leqq j \leqq n$. Then, using the distributive property of matric multiplication and the associative and commutative properties of matric addition, we have

$$\begin{aligned} f(A)g(A) &= \left(\sum_{i=0}^{m} a_i A^i\right)\left(\sum_{j=0}^{n} b_j A^j\right) \\ &= \sum_{i=0}^{m}\sum_{j=0}^{n} (a_i A^i)(b_j A^j) = \sum_{i=0}^{m}\sum_{j=0}^{n} a_i b_j A^{i+j} \end{aligned}$$

or, grouping together all terms involving the same power of A,

$$\begin{aligned} f(A)g(A) &= \sum_{k=0}^{t=m+n} c_k A^k \text{ where } k = i + j \text{ and } c_k = \sum_{i,j} a_i b_j \\ &= p(A) \qquad \text{as asserted} \end{aligned}$$

Since $f(x)g(x) = g(x)f(x)$, it follows from Theorem 1 that

(2) $$f(A)g(A) = g(A)f(A)$$

In other words, we have the following important result.

COROLLARY 1

Any two polynomials in a matrix A commute with each other.

If $g(A)$ is a nonsingular matrix, then $g^{-1}(A)$ exists, and we may premultiply and postmultiply each side of Eq. (2) by $g^{-1}(A)$, getting

$$g^{-1}(A)f(A)g(A)g^{-1}(A) = g^{-1}(A)g(A)f(A)g^{-1}(A)$$

or

$$g^{-1}(A)f(A) = f(A)g^{-1}(A) \tag{3}$$

With this identity, we are now in a position to define rational fractional functions of a square matrix A

DEFINITION 2

If $f(x)$ and $g(x)$ are scalar polynomials and if A is a square matrix such that $g(A)$ is nonsingular, then either of the equal matrices $g^{-1}(A)f(A)$ and $f(A)g^{-1}(A)$ is called the quotient of $f(A)$ by $g(A)$ and is written $f(A)/g(A)$.

It is now relatively easy to prove the following extension of Theorem 1 (see Exercise 3).

THEOREM 2

Any identity between rational fractional functions of a scalar variable implies a corresponding matric identity, provided all the matric functions are defined

With rational functions of a square matrix now defined, it is natural to ask whether the characteristic values of a rational function of a matrix A can be expressed in terms of the characteristic values of A. This is indeed the case, as the following chain of theorems makes clear.

THEOREM 3

If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the (possibly repeated) characteristic values of a square matrix A and if f is any polynomial, then

$$|f(A)| = f(\lambda_1)f(\lambda_2) \cdots f(\lambda_n)$$

PROOF Let the characteristic polynomial of the given matrix A be

$$|A - \lambda I| = \prod_{i=1}^{n} (\lambda_i - \lambda) \tag{4}$$

and let the factored form of the given polynomial f be

$$f(t) = c(t - r_1)(t - r_2) \cdots (t - r_k) \tag{5}$$

Then, since Theorem 1 assures us that identities between scalar polynomials imply corresponding matric identities, we have

$$f(A) = c(A - r_1I)(A - r_2I) \cdots (A - r_kI)$$

Furthermore, since the determinant of a product of square matrices is equal to the product of the determinants of the matric factors, and since the scalar factor c

incorporated into any one of the matric factors reappears as the factor c^n in the determinant of that matrix, we have

$$(6) \qquad |f(A)| = c^n|A - r_1I|\,|A - r_2I| \cdots |A - r_kI| = c^n \prod_{j=1}^{k} |A - r_jI|$$

However, $|A - r_jI|$ is just the characteristic polynomial of A evaluated for $\lambda = r_j$. Hence, by (4),

$$|A - r_jI| = \prod_{i=1}^{n} (\lambda_i - r_j)$$

and, therefore, substituting into (6), we have

$$|f(A)| = c^n \prod_{j=1}^{k} \prod_{i=1}^{n} (\lambda_i - r_j)$$

Next, interchanging the order in which the products are formed by first grouping together all the factors corresponding to a given value of i and assigning a single factor c to each such group, we have

$$|f(A)| = c^n \prod_{i=1}^{n} \prod_{j=1}^{k} (\lambda_i - r_j) = \prod_{i=1}^{n} \left[c \prod_{j=1}^{k} (\lambda_i - r_j)\right]$$

Finally we observe that, with the coefficient c, the inner product in the last expression is precisely the evaluation of the factored form (5) of the given polynomial for $t = \lambda_i$. Hence,

$$|f(A)| = \prod_{i=1}^{n} f(\lambda_i)$$

as asserted.

THEOREM 4

If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the characteristic values of a square matrix A, if $f = g/h$ is a rational fractional function, and if $|h(A)|$ is different from zero, then

$$|f(A)| = f(\lambda_1)f(\lambda_2) \cdots f(\lambda_n)$$

PROOF Since, by definition, $f(A) = g(A)/h(A) = g(A)h^{-1}(A)$ and since the determinant of a product of square matrices is equal to the product of the determinants of the matric factors, we have

$$|f(A)| = |g(A)h^{-1}(A)| = |g(A)|\,|h^{-1}(A)|$$

Moreover, as we observed in Sec. 10.3, $|h^{-1}(A)| = 1/|h(A)|$. Therefore,

$$|f(A)| = \frac{|g(A)|}{|h(A)|}$$

However, by Theorem 3, since g and h are polynomials,

$$|g(A)| = g(\lambda_1)g(\lambda_2) \cdots g(\lambda_n) \quad \text{and} \quad |h(A)| = h(\lambda_1)h(\lambda_2) \cdots h(\lambda_n)$$

Hence,

$$|f(A)| = \frac{g(\lambda_1)g(\lambda_2) \cdots g(\lambda_n)}{h(\lambda_1)h(\lambda_2) \cdots h(\lambda_n)} = f(\lambda_1)f(\lambda_2) \cdots f(\lambda_n)$$

as asserted

THEOREM 5

If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the characteristic values of a square matrix A and if $f = g/h$, where g and h are polynomials such that $|h(A)| \neq 0$, then the characteristic values of $f(A)$ are $f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)$.

PROOF Let

$$\phi(x) = f(x) - \lambda = \frac{g(x)}{h(x)} - \lambda = \frac{g(x) - \lambda h(x)}{h(x)}$$

Clearly, $g(x) - \lambda h(x)$ is a polynomial, and, therefore, $\phi(x)$ is a rational fractional function of x. Hence, by the last theorem,

$$|\phi(A)| = \phi(\lambda_1)\phi(\lambda_2) \cdots \phi(\lambda_n)$$

In other words, for all values of λ,

$$|f(A) - \lambda I| = [f(\lambda_1) - \lambda][f(\lambda_2) - \lambda] \cdots [f(\lambda_n) - \lambda]$$

The right-hand side of this identity is, thus, the factored form of the characteristic polynomial $|f(A) - \lambda I|$ of the matrix $f(A)$; hence, the roots of the characteristic equation of $f(A)$ are

$$\lambda = f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n) \qquad \text{as asserted}$$

COROLLARY 1

If the characteristic values of a matrix A are $\lambda_1, \lambda_2, \ldots, \lambda_n$, then for all integral values of k if A is nonsingular and for all nonnegative integral values of k if A is singular, the characteristic values of A^k are $\lambda_1{}^k, \lambda_2{}^k, \ldots, \lambda_n{}^k$

COROLLARY 2

If X_i is a characteristic vector corresponding to the characteristic value λ_i of a square matrix A and if p is a polynomial, then X_i is also a characteristic vector corresponding to the characteristic value $p(\lambda_i)$ of the matrix $p(A)$.

EXAMPLE 2

As an illustration of Theorem 5, consider the matrix $A = \begin{Vmatrix} 1 & -2 \\ 3 & -4 \end{Vmatrix}$ and the function $\phi(x) = x/(x + 3)$ The characteristic equation of A is

$$|A - \lambda I| = \begin{vmatrix} 1 - \lambda & -2 \\ 3 & -4 - \lambda \end{vmatrix} = \lambda^2 + 3\lambda + 2 = 0$$

Hence, the characteristic roots are $\lambda = -1, -2$ Therefore, according to Theorem 5, the characteristic roots of $\phi(A)$ are

$$\phi(-1) = -\tfrac{1}{2} \qquad \text{and} \qquad \phi(-2) = -2$$

To confirm this, we have, by direct calculation,

$$\begin{aligned} \phi(A) = \frac{A}{A + 3I} = A(A + 3I)^{-1} &= \begin{Vmatrix} 1 & -2 \\ 3 & -4 \end{Vmatrix} \cdot \begin{Vmatrix} 4 & -2 \\ 3 & -1 \end{Vmatrix}^{-1} \\ &= \begin{Vmatrix} 1 & -2 \\ 3 & -4 \end{Vmatrix} \tfrac{1}{2} \begin{Vmatrix} -1 & 2 \\ -3 & 4 \end{Vmatrix} \\ &= \tfrac{1}{2} \begin{Vmatrix} 5 & -6 \\ 9 & -10 \end{Vmatrix} = \begin{Vmatrix} \tfrac{5}{2} & -3 \\ \tfrac{9}{2} & -5 \end{Vmatrix} \end{aligned}$$

The characteristic roots of $\phi(A)$ are, therefore, the roots of the equation

$$|\phi(A) - \lambda I| = \begin{vmatrix} \frac{5}{2} - \lambda & -3 \\ \frac{9}{2} & -5 - \lambda \end{vmatrix} = \lambda^2 + \tfrac{5}{2}\lambda + 1 = 0$$

or $-\frac{1}{2}$ and -2, as before

If p is a polynomial and A is a square matrix, the evaluation of $p(A)$ is a perfectly straightforward matter However, when A is a matrix similar to a diagonal matrix, the evaluation of $p(A)$ can be appreciably simplified. To establish the result upon which this simplification is based, it is convenient first to prove the following lemmas

LEMMA 1

If $B = S^{-1}AS$, then $B^n = S^{-1}A^nS$

PROOF Clearly, the lemma is true for $n = 2$, since

$$B^2 = (S^{-1}AS)(S^{-1}AS) = S^{-1}A(SS^{-1})AS = S^{-1}A^2S$$

Assuming, then, that the lemma is true for $n = k$, we have

$$B^{k+1} = BB^k = (S^{-1}AS)(S^{-1}A^kS) = S^{-1}A(SS^{-1})A^kS = S^{-1}A^kS$$

which completes the induction and establishes the lemma

If we now apply Lemma 1 to each term of any polynomial function of B and then use the distributive property of matric multiplication, we obtain the following result·

LEMMA 2

If $B = S^{-1}AS$ and if p is a polynomial, then $p(B) = S^{-1}p(A)S$, i e.,

$$p(S^{-1}AS) = S^{-1}p(A)S$$

Furthermore, by another easy induction we can establish the following observation

LEMMA 3

If D is the diagonal matrix

$$\begin{Vmatrix} d_{11} & & \bigcirc \\ & d_{22} & \\ & & \ddots & \\ \bigcirc & & & d_{nn} \end{Vmatrix} \qquad \text{then} \qquad D^k = \begin{Vmatrix} d_{11}{}^k & & \bigcirc \\ & d_{22}{}^k & \\ & & \ddots & \\ \bigcirc & & & d_{nn}{}^k \end{Vmatrix}$$

Finally, by applying Lemma 3 to each term of any polynomial function of a diagonal matrix D and then using the definition of matric addition, we have the following result:

LEMMA 4

If D is the diagonal matrix

$$\begin{Vmatrix} d_{11} & & \bigcirc \\ & d_{22} & \\ & & \ddots & \\ \bigcirc & & & d_{nn} \end{Vmatrix}$$

and p is any polynomial, then

$$p(D) = \begin{Vmatrix} p(d_{11}) & & \bigcirc \\ & p(d_{22}) & \\ \bigcirc & & p(d_{nn}) \end{Vmatrix}$$

Using Lemmas 2 and 4, we can now prove the following useful theorem·

THEOREM 6

If a matrix A is similar to a diagonal matrix; i e , if

$$S^{-1}AS = D = \begin{Vmatrix} \lambda_1 & & \bigcirc \\ & \lambda_2 & \\ & & \cdot \\ \bigcirc & & \lambda_n \end{Vmatrix}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the characteristic values of A, then

$$p(A) = S \begin{Vmatrix} p(\lambda_1) & & \bigcirc \\ & p(\lambda_2) & \\ \bigcirc & & p(\lambda_n) \end{Vmatrix} S^{-1}$$

PROOF By Lemma 4,

$$p(D) = \begin{Vmatrix} p(\lambda_1) & & \bigcirc \\ & p(\lambda_2) & \\ \bigcirc & & p(\lambda_n) \end{Vmatrix}$$

Also, since $S^{-1}AS = D$, it follows that $A = SDS^{-1}$ Hence, using Lemma 2, we have

$$S \begin{Vmatrix} p(\lambda_1) & & \bigcirc \\ & p(\lambda_2) & \\ & & \cdot \\ \bigcirc & & p(\lambda_n) \end{Vmatrix} S^{-1} = Sp(D)S^{-1} = p(SDS^{-1}) = p(A) \quad \text{as asserted}$$

EXAMPLE 3

If $p(x) = x^4 - 4x^3 + 6x^2 - x - 3$ and $A = \begin{Vmatrix} 0 & -2 \\ 1 & 3 \end{Vmatrix}$, what is $p(A)$?

By an easy calculation we find the characteristic equation of A to be

$$|A - \lambda I| = \begin{vmatrix} -\lambda & -2 \\ 1 & 3 - \lambda \end{vmatrix} = \lambda^2 - 3\lambda + 2 = 0$$

Hence, the characteristic values of A are $\lambda_1 = 1$ and $\lambda_2 = 2$; and, since these are distinct, it follows from Theorem 4, Sec 11 3, that A is similar to a diagonal matrix and Theorem 6 can be applied Now, corresponding to λ_1 and λ_2 we have the characteristic vectors

$$X_1 = \begin{Vmatrix} 2 \\ -1 \end{Vmatrix} \quad \text{and} \quad X_2 = \begin{Vmatrix} 1 \\ -1 \end{Vmatrix}$$

and from these we can construct the modal matrix

$$S = \begin{Vmatrix} 2 & 1 \\ -1 & -1 \end{Vmatrix} \quad \text{and its inverse} \quad S^{-1} = \begin{Vmatrix} 1 & 1 \\ -1 & -2 \end{Vmatrix}$$

According to Theorem 4, Sec. 11.3, these are matrices such that

$$S^{-1}AS = D = \begin{Vmatrix} 1 & 0 \\ 0 & 2 \end{Vmatrix}$$

Hence these are the matrices to be used in evaluating $p(A)$ by means of Theorem 6. Now,

$$p(\lambda_1) = p(1) = -1 \qquad \text{and} \qquad p(\lambda_2) = p(2) = 3$$

Therefore,

$$\begin{aligned} p(A) = A^4 - 4A^3 + 6A^2 - A - 3I &= S \begin{Vmatrix} p(\lambda_1) & 0 \\ 0 & p(\lambda_2) \end{Vmatrix} S^{-1} \\ &= \begin{Vmatrix} 2 & 1 \\ -1 & -1 \end{Vmatrix} \begin{Vmatrix} -1 & 0 \\ 0 & 3 \end{Vmatrix} \begin{Vmatrix} 1 & 1 \\ -1 & -2 \end{Vmatrix} \\ &= \begin{Vmatrix} -5 & -8 \\ 4 & 7 \end{Vmatrix} \end{aligned}$$

After polynomial functions of a square matrix have been defined, it is natural to consider polynomial equations in a matric variable. In particular, now that we have developed procedures for evaluating $p(A)$, that is, solving the equation $p(A) = X$, we shall consider the problem of solving the nontrivial equation $p(X) = A$, where p is a given polynomial, A is a given square matrix, and X is a matric variable. By means of examples (see Exercise 1) it is easy to show that there are polynomial equations $p(X) = A$ which have no solution. In one important case, however, the equation $p(X) = A$ can always be solved, as the following theorem makes clear:

THEOREM 7

If A is similar to a diagonal matrix and if p is a scalar polynomial, the equation $p(X) = A$ is solvable for X.

PROOF By hypothesis, since A is similar to a diagonal matrix D, there exists a nonsingular matrix S with the property that $S^{-1}AS = D$, or $A = SDS^{-1}$, where, say,

$$D = \begin{Vmatrix} d_{11} & & \bigcirc \\ & d_{22} & \\ \bigcirc & & d_{nn} \end{Vmatrix}$$

Now, let r_i be one of the roots of the equation $p(x) = d_{ii}$. Then, if

$$R = \begin{Vmatrix} r_1 & & \bigcirc \\ & r_2 & \\ \bigcirc & & r_n \end{Vmatrix} \qquad \text{and} \qquad X = SRS^{-1}$$

we have, by Lemma 2, noting that $S = (S^{-1})^{-1}$,

$$p(X) = p(SRS^{-1}) = Sp(R)S^{-1}$$

Moreover, by Lemma 4 and the fact that $p(r_i) = d_{ii}$,

$$p(R) = \begin{Vmatrix} p(r_1) & & \bigcirc \\ & p(r_2) & \\ & & \ddots \\ \bigcirc & & p(r_n) \end{Vmatrix} = \begin{Vmatrix} d_{11} & & \bigcirc \\ & d_{22} & \\ & & \ddots \\ \bigcirc & & d_{nn} \end{Vmatrix} = D$$

Therefore, $\quad p(X) = Sp(R)S^{-1} = SDS^{-1} = A$

which proves that, if A is similar to a diagonal matrix, then $p(X) = A$ has the solution $X = SRS^{-1}$. If the polynomial p is of degree k, the scalar equation $p(x) = d_{ii}$ has, in general, k distinct roots. Hence, there are k distinct choices for each of the n diagonal elements in R, and, therefore, $p(X) = A$ has, in general, at least k^n different solutions.

By applying the preceding theorem to the particular equation $X^2 = A$, we obtain the following corollary:

COROLLARY 1

An (n,n) matrix with distinct characteristic values has at least 2^n or 2^{n-1} distinct square roots, according as it is nonsingular or singular

PROOF Let A be an (n,n) matrix with n distinct characteristic values $\lambda_1, \lambda_2, \ldots, \lambda_n$. It follows, then, by Theorem 4, Sec 11.3, that there exists a nonsingular matrix S such that

$$A = S \begin{Vmatrix} \lambda_1 & & \bigcirc \\ & \lambda_2 & \\ & & \ddots \\ \bigcirc & & \lambda_n \end{Vmatrix} S^{-1}$$

Thus, according to the last theorem, for any choice of plus and minus signs,

$$X = S \begin{Vmatrix} \pm\sqrt{\lambda_1} & & \bigcirc \\ & \pm\sqrt{\lambda_2} & \\ \bigcirc & & \pm\sqrt{\lambda_n} \end{Vmatrix} S^{-1}$$

satisfies the equation $X^2 = A$ If A is nonsingular, none of the λ's is zero, and there are 2^n combinations of signs each leading to a different matrix X satisfying the equation $X^2 = A$ On the other hand, if A is singular but still has distinct characteristic values, then, by Theorem 5, Sec 11.2, one and only one of the λ's must be zero, and, therefore, for one of the diagonal elements there is only a single choice rather than two Hence, in this case there may be no more than 2^{n-1} distinct square roots, as asserted.

EXAMPLE 4

Solve the equation $X^2 - 4X + 4I = \begin{Vmatrix} 4 & 3 \\ 5 & 6 \end{Vmatrix}$

The characteristic equation of the matrix $A = \begin{Vmatrix} 4 & 3 \\ 5 & 6 \end{Vmatrix}$ is

$$\begin{vmatrix} 4-\lambda & 3 \\ 5 & 6-\lambda \end{vmatrix} = \lambda^2 - 10\lambda + 9 = 0$$

Hence, the characteristic values of A are $\lambda_1 = 1$, $\lambda_2 = 9$, and the corresponding characteristic vectors are $X_1 = \begin{Vmatrix} 1 \\ -1 \end{Vmatrix}$, $X_2 = \begin{Vmatrix} 3 \\ 5 \end{Vmatrix}$. Therefore, by Theorem 4, Sec. 11.3, A is similar to a diagonal matrix; that is,

$$S^{-1}AS = D \qquad \text{or} \qquad A = SDS^{-1}$$

where S is the modal matrix $\begin{Vmatrix} 1 & 3 \\ -1 & 5 \end{Vmatrix}$, $S^{-1} = \frac{1}{8}\begin{Vmatrix} 5 & -3 \\ 1 & 1 \end{Vmatrix}$, and $D = \begin{Vmatrix} 1 & 0 \\ 0 & 9 \end{Vmatrix}$. We must now solve the equations $p(x) = d_{ii}$ for r_i $(i = 1,2)$.

$$x^2 - 4x + 4 = d_{11} = 1 \qquad x^2 - 4x + 4 = d_{22} = 9$$
$$x = r_1 = 1, 3 \qquad x = r_2 = -1, 5$$

Pairing each possibility for r_1 with each possibility for r_2, we thus obtain four possibilities for the matrix R:

$$R_1 = \begin{Vmatrix} 1 & 0 \\ 0 & -1 \end{Vmatrix} \qquad R_2 = \begin{Vmatrix} 1 & 0 \\ 0 & 5 \end{Vmatrix} \qquad R_3 = \begin{Vmatrix} 3 & 0 \\ 0 & -1 \end{Vmatrix} \qquad R_4 = \begin{Vmatrix} 3 & 0 \\ 0 & 5 \end{Vmatrix}$$

Then, according to Theorem 7, the solutions of the given equation are

$$X_1 = SR_1S^{-1} = \begin{Vmatrix} 1 & 3 \\ -1 & 5 \end{Vmatrix} \begin{Vmatrix} 1 & 0 \\ 0 & -1 \end{Vmatrix} \frac{1}{8}\begin{Vmatrix} 5 & -3 \\ 1 & 1 \end{Vmatrix} = \frac{1}{4}\begin{Vmatrix} 1 & -3 \\ -5 & -1 \end{Vmatrix}$$

and, similarly,

$$X_2 = SR_2S^{-1} = \frac{1}{2}\begin{Vmatrix} 5 & 3 \\ 5 & 7 \end{Vmatrix} \qquad X_3 = SR_3S^{-1} = \frac{1}{2}\begin{Vmatrix} 3 & -3 \\ -5 & 1 \end{Vmatrix}$$

$$X_4 = SR_4S^{-1} = \frac{1}{4}\begin{Vmatrix} 15 & 3 \\ 5 & 17 \end{Vmatrix}$$

EXERCISES

1 Prove that there is no matrix which satisfies the equation $X^2 = \begin{Vmatrix} 0 & 1 \\ 0 & 0 \end{Vmatrix}$.

2 Show that, for particular polynomials and particular matrices, each of the following cases is possible:

a A nonsingular, $p(A)$ nonsingular

b A nonsingular, $p(A)$ singular

c A singular, $p(A)$ nonsingular

d A singular, $p(A)$ singular

3 Prove Theorem 2. [Hint: Note first that it is sufficient to prove that

$$\frac{f_1(x)}{f_2(x)} + \frac{f_3(x)}{f_4(x)} = \frac{f_5(x)}{f_6(x)} \quad \text{implies} \quad \frac{f_1(A)}{f_2(A)} + \frac{f_3(A)}{f_4(A)} = \frac{f_5(A)}{f_6(A)}$$

$$\frac{f_1(x)}{f_2(x)} \frac{f_3(x)}{f_4(x)} = \frac{f_5(x)}{f_6(x)} \quad \text{implies} \quad \frac{f_1(A)}{f_2(A)} \frac{f_3(A)}{f_4(A)} = \frac{f_5(A)}{f_6(A)}$$

$$\frac{f_1(x)/f_2(x)}{f_3(x)/f_4(x)} = \frac{f_5(x)}{f_6(x)} \quad \text{implies} \quad \frac{f_1(A)/f_2(A)}{f_3(A)/f_4(A)} = \frac{f_5(A)}{f_6(A)}$$

Then clear of fractions in the scalar identities, use Theorem 1, and multiply the resulting matric identities by the appropriate inverses.]

4 If A is a diagonal matrix and p is any scalar polynomial, show that $p(A)$ is also a diagonal matrix.

5 If A is a diagonal matrix and f is a rational fractional function, is $f(A)$ necessarily a diagonal matrix?

6 Show that I_2 has infinitely many distinct square roots.

7 Prove that an (n,n) matrix with distinct characteristic values has no square roots other than those identified by Corollary 1, Theorem 7. (Hint: Use the result of Exercise 6, Sec. 10.2.)

8 Prove Corollary 2, Theorem 5. [Hint: First prove the assertion for the special polynomials $p(A) = A^k$ by premultiplying the equation $AX_i = \lambda_i X_i$ by $A, A^2, \ldots, A^{k-1}$, in turn.]

9 By actually constructing an infinite family of solutions, show that each of the following matric equations is satisfied by infinitely many matrices:

a $X^2 - 2X - 3I_2 = O$ b $X^2 - 4X + 3I_2 = O$
c $X^2 - 4X - 5I_2 = O$ d $X^3 - 6X^2 + 11X - 6I_2 = O$

10 Without attempting to find the solutions, show that, for all values of a and b, the equation $X^2 + aX + bI_2 = O$ is satisfied by infinitely many matrices What do you think is the generalization of this result to equations in an (n,n) matric variable?

11 Show that the following matric equations have no solutions:

a $X^2 - 2X - 3I_2 = \begin{Vmatrix} -4 & 1 \\ 0 & -4 \end{Vmatrix}$ b $X^2 - 4X + 3I_2 = \begin{Vmatrix} -1 & 2 \\ 0 & -1 \end{Vmatrix}$

c $X^2 - 4X - 5I_2 = \begin{Vmatrix} -9 & 3 \\ 0 & -9 \end{Vmatrix}$ d $X^2 - 4X + 3I_3 = \begin{Vmatrix} 2 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{Vmatrix}$

12 If A and B commute, show that A commutes with any polynomial in B

13 Verify each of the following identities for $X = \begin{Vmatrix} 1 & 2 \\ 1 & 3 \end{Vmatrix}$ and $X = \begin{Vmatrix} 0 & 2 \\ 0 & -2 \end{Vmatrix}$

a $(X - I)^2 = X^2 - 2X + I$ b $X^3 - I = (X - I)(X^2 + X + I)$

c $\dfrac{2X}{X^2 - I} = \dfrac{I}{X - I} + \dfrac{I}{X + I}$ d $\dfrac{X^2}{X - 2I} = X + 2I + \dfrac{4I}{X - 2I}$

14 If $A^k = O$ for some positive integer k, prove that every characteristic value of A is zero

15 Solve each of the following matric equations:

a $X^2 - 5X + 3I = \begin{Vmatrix} 1 & -4 \\ 2 & -5 \end{Vmatrix}$ b $X^2 + 6X + 9I = \begin{Vmatrix} -5 & 9 \\ -6 & 10 \end{Vmatrix}$

c $X^2 = \begin{Vmatrix} -6 & 14 \\ -7 & 15 \end{Vmatrix}$ d $X^2 = \begin{Vmatrix} 8 & -7 & -7 \\ -9 & 10 & 11 \\ 9 & -9 & -10 \end{Vmatrix}$

16 If $f(x) = x/(x + 4)$, compute $f(A)$ for each of the following matrices A

a $\begin{Vmatrix} 1 & -4 \\ 2 & -5 \end{Vmatrix}$ b $\begin{Vmatrix} 4 & -1 \\ 6 & -1 \end{Vmatrix}$ c $\begin{Vmatrix} 2 & -1 \\ 4 & -3 \end{Vmatrix}$

d $\begin{Vmatrix} -3 & 1 & 0 \\ 1 & -4 & 0 \\ 0 & 0 & -5 \end{Vmatrix}$ e $\begin{Vmatrix} -1 & 2 & 2 \\ 2 & -1 & -2 \\ -2 & 2 & 3 \end{Vmatrix}$

17 If $p(x) = x^4 - x^3 - 3x^2 + 4x + 2$, evaluate $p(A)$ for each of the following matrices A·

a $\begin{Vmatrix} 4 & 1 \\ -3 & 0 \end{Vmatrix}$ b $\begin{Vmatrix} -1 & -2 \\ 3 & 4 \end{Vmatrix}$ c $\begin{Vmatrix} 4 & 6 \\ -3 & -5 \end{Vmatrix}$

d $\begin{Vmatrix} -4 & -9 & -3 \\ 1 & 4 & 1 \\ 3 & 3 & 2 \end{Vmatrix}$ e $\begin{Vmatrix} 2 & 1 & 1 \\ 1 & 4 & 3 \\ -1 & -1 & 0 \end{Vmatrix}$

18 Verify that the characteristic values of $f(A)$ are equal to $f(\lambda_i)$ for each of the following functions and each of the given matrices·

a $x^2 - 2x + 3$ b $x^2 - 4x + 3$ c $x^3 + x^2 + x + 1$

i $\begin{Vmatrix} -1 & -4 \\ 2 & -5 \end{Vmatrix}$ ii $\begin{Vmatrix} 4 & -1 \\ 6 & -1 \end{Vmatrix}$ iii $\begin{Vmatrix} 2 & 1 \\ 1 & 2 \end{Vmatrix}$

19 Verify that the characteristic values of $f(A)$ are equal to $f(\lambda_i)$ for each of the following functions and each of the given matrices.

a $\dfrac{x}{x^2+1}$ b $\dfrac{x-2}{x+2}$ c $\dfrac{x+1}{x^2+x+1}$

i $\begin{Vmatrix} -1 & 2 \\ -1 & 2 \end{Vmatrix}$ ii $\begin{Vmatrix} 4 & 2 \\ 1 & 3 \end{Vmatrix}$ iii $\begin{Vmatrix} 5 & 2 \\ 2 & 2 \end{Vmatrix}$

20 If p is a polynomial, is it possible for $p(A)$ to have a characteristic vector which is not a characteristic vector of A? Justify your answer.

11.5

The Cayley-Hamilton theorem

Since a square null matrix, being a diagonal matrix, is obviously similar to a diagonal matrix, it follows from Theorem 7, Sec. 11.4, that the equation $p(X) = O$ is always solvable. On the other hand, it is not immediately evident that, given a square matrix A, there is always a polynomial equation with scalar coefficients $p(X) = O$ of which A is a solution. This is the case, however, and it is not difficult to show (see Exercise 1) that any square matrix of order n satisfies a polynomial equation whose order is at most n^2. In fact, for any square matrix A, there is always a polynomial equation of order n which is satisfied by A.

To prove this, it is convenient to prove first a preliminary result concerning polynomials whose coefficients are not scalars but square matrices. Before we can do this, however, it is necessary that we define what is meant by the value of such a polynomial, say

$$F(\lambda) = C_0 + C_1\lambda + \cdots + C_k\lambda^k$$

when a square matrix A is substituted for the scalar variable λ. Since matric multiplication is not commutative, it is clear that, in general, the various powers of A will not commute with the coefficient matrices in $F(\lambda)$. Hence, although it is true that

$$C_0 + C_1\lambda + \cdots + C_k\lambda^k = C_0 + \lambda C_1 + \cdots + \lambda^k C_k$$

the corresponding matric relation, namely,

$$C_0 + C_1A + \cdots + C_kA^k = C_0 + AC_1 + \cdots + A^kC_k$$

is in general false. Thus it is necessary for us to assign a meaning to $F(A)$, and this we do by agreeing that

$$F(A) = C_0 + C_1A + \cdots + C_kA^k$$

Now we have already seen (Theorem 1, Sec. 11.4) that identities relating scalar polynomials imply corresponding identities when the scalar variable is replaced by a square matrix.

This is not true, however, for identical relations involving polynomials with matric coefficients. For instance, if

$$F(\lambda) = C_0 + C_1\lambda \qquad \text{and} \qquad G(\lambda) = D_0 + D_1\lambda$$

then, for the product $F(\lambda)G(\lambda)$, we have

$$P(\lambda) = C_0D_0 + (C_0D_1 + C_1D_0)\lambda + C_1D_1\lambda^2$$

On the other hand, if we replace the scalar λ by a square matrix A, we have

$$F(A) = C_0 + C_1A \qquad G(A) = D_0 + D_1A$$

and $$F(A)G(A) = C_0D_0 + C_0D_1A + C_1AD_0 + C_1AD_1A$$

which is not equal to $P(A)$ unless

$$AD_0 = D_0A \qquad \text{and} \qquad AD_1 = D_1A$$

However, we can prove the following theorem, which is the necessary preliminary result we mentioned above

THEOREM 1

If $F(\lambda)$ and $P(\lambda)$ are polynomials in the scalar variable λ with coefficients which are square matrices and if $P(\lambda) = F(\lambda)(A - \lambda I)$, then $P(A) = O$

PROOF In view of the fact that we have just seen that $P(\lambda) = F(\lambda)G(\lambda)$ does not imply that $P(A) = F(A)G(A)$, we cannot prove this theorem simply by substituting A for λ in the assertion of the theorem Instead we must first multiply out the right-hand side of the given relation, express it as a polynomial in λ, and then replace λ by A. To do this, let us suppose that

$$F(\lambda) = C_0 + C_1\lambda + C_2\lambda^2 + \cdots + C_k\lambda^k$$

where $C_0, C_1, C_2, \ldots, C_k$ are (n,n) matrices Then

$$\begin{aligned} P(\lambda) &= (C_0 + C_1\lambda + C_2\lambda^2 + \cdots + C_k\lambda^k)(A - \lambda I) \\ &= C_0A + C_1A\lambda + C_2A\lambda^2 + \cdots + C_kA\lambda^k \\ &\qquad - C_0\lambda - C_1\lambda^2 - \cdots - C_{k-1}\lambda^k - C_k\lambda^{k+1} \\ &= C_0A + (C_1A - C_0)\lambda + (C_2A - C_1)\lambda^2 + \cdots \\ &\qquad + (C_kA - C_{k-1})\lambda^k - C_k\lambda^{k+1} \end{aligned}$$

Now, substituting A for λ, we have

$$\begin{aligned} P(A) &= C_0A + (C_1A - C_0)A + (C_2A - C_1)A^2 + \cdots \\ &\qquad + (C_kA - C_{k-1})A^k - C_kA^{k+1} \\ &= O \qquad \text{as asserted.} \end{aligned}$$

We are now in a position to prove one of the most important results in the theory of matrices, the famous **Cayley-Hamilton theorem:**

THEOREM 2

Every square matrix satisfies its own characteristic equation.

PROOF Let A be an (n,n) matrix whose characteristic equation is

$$|A - \lambda I| = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdot & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdot\ \cdot & a_{2n} \\ \cdot\ \cdot & \cdot\ \cdot & \cdot\ \cdot & \cdot\ \cdot \\ a_{n1} & a_{n2} & \cdot & a_{nn} - \lambda \end{vmatrix}$$

$$= (-1)^n[\lambda^n - \beta_1\lambda^{n-1} + \cdots + (-1)^n\beta_n] = 0$$

The adjoint of the matrix $A - \lambda I$ is clearly an (n,n) matrix whose elements, being the cofactors of the elements of the determinant $|A - \lambda I|$, are polynomials in λ; that is,

$$\operatorname{adj}(A - \lambda I) = \begin{Vmatrix} p_{11}(\lambda) & p_{12}(\lambda) & \cdot & p_{1n}(\lambda) \\ p_{21}(\lambda) & p_{22}(\lambda) & & p_{2n}(\lambda) \\ & & \cdot & \cdot\ \cdot \\ p_{n1}(\lambda) & p_{n2}(\lambda) & & p_{nn}(\lambda) \end{Vmatrix}$$

Furthermore, from the definition of matric addition, it follows that the last matrix can be written as a polynomial in λ, say $F(\lambda)$, whose coefficients are (n,n) matrices, the element in the ith row and jth column of the matric coefficient of λ^k being the coefficient of λ^k in $p_{ij}(\lambda)$. Now, from Corollary 1, Theorem 1, Sec. 10.3, we have

$$\begin{aligned} \operatorname{adj}(A - \lambda I)\,(A - \lambda I) &= |A - \lambda I|I \\ &= (-1)^n[\lambda^n I - \beta_1\lambda^{n-1}I + \cdots + (-1)^n\beta_n I] \end{aligned}$$

that is,

$$(-1)^n[\lambda^n I - \beta_1\lambda^{n-1}I + \cdots + (-1)^n\beta_n I] = F(\lambda)(A - \lambda I)$$

But this is a relation between polynomials in λ with matric coefficients of precisely the type covered by Theorem 1. Hence, the left-hand side must vanish when $\lambda = A$. In other words,

$$A^n - \beta_1 A^{n-1} + \cdots + (-1)^n\beta_n I = O$$

that is, the matrix A satisfies its own characteristic equation, as asserted.

Using the Cayley-Hamilton theorem, the nth power of any square matrix A can be expressed as a linear combination of lower powers of A. Hence, by repeated applications of the Cayley-Hamilton theorem, any positive integral power of A and, therefore, any polynomial in A can be expressed as a polynomial in A of degree at most $n - 1$. Moreover, if A is nonsingular, then A^{-1} exists, and, in the expansion of $|A - \lambda I|$, the constant term $\beta_n = |A|$ is different from zero. Hence, we can multiply the Cayley-Hamilton equation

$$A^n - \beta_1 A^{n-1} + \cdots + (-1)^n\beta_n I = O$$

by A^{-1}, getting

$$A^{n-1} - \beta_1 A^{n-2} + \cdots + (-1)^{n-1}\beta_{n-1}I + (-1)^n\beta_n A^{-1} = O$$

whence, solving for A^{-1}, we find

$$A^{-1} = \frac{(-1)^{n-1}}{\beta_n}[A^{n-1} - \beta_1 A^{n-2} + \cdots + (-1)^{n-1}\beta_{n-1}I]$$

In some cases this is a convenient method of obtaining the inverse of a matrix A

EXAMPLE 1

If $A = \begin{Vmatrix} -4 & 5 & 5 \\ -5 & 6 & 5 \\ -5 & 5 & 6 \end{Vmatrix}$ we find by an easy calculation that

$$|A - \lambda I| = -\lambda^3 + 8\lambda^2 - 13\lambda + 6$$

Hence, by the Cayley-Hamilton theorem it follows that

$$A^3 - 8A^2 + 13A - 6I = O$$

as can easily be verified by direct calculation Using this relation, we can now express higher powers of A as quadratic polynomials in A For instance,

$$\begin{aligned} A^4 = A\ A^3 &= A(8A^2 - 13A + 6I) \\ &= 8A^3 - 13A^2 + 6A \\ &= 8(8A^2 - 13A + 6I) - 13A^2 + 6A \\ &= 51A^2 - 98A + 48I \end{aligned}$$

and

$$\begin{aligned} A^5 = A\ A^4 &= A(51A^2 - 98A + 48I) \\ &= 51(8A^2 - 13A + 6I) - 98A^2 + 48A \\ &= 310A^2 - 615A + 306I \end{aligned}$$

Similarly, multiplying the Cayley-Hamilton equation through by A^{-1} and then solving for A^{-1}, we find

$$\begin{aligned} A^{-1} &= \tfrac{1}{6}(A^2 - 8A + 13I) \\ &= \frac{1}{6}\left(\begin{Vmatrix} -34 & 35 & 35 \\ -35 & 36 & 35 \\ -35 & 35 & 36 \end{Vmatrix} - 8\begin{Vmatrix} -4 & 5 & 5 \\ -5 & 6 & 5 \\ -5 & 5 & 6 \end{Vmatrix} + 13\begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}\right) \\ &= \frac{1}{6}\begin{Vmatrix} 11 & -5 & -5 \\ 5 & 1 & -5 \\ 5 & -5 & 1 \end{Vmatrix} \end{aligned}$$

The Cayley-Hamilton equation is not necessarily the polynomial equation of lowest degree satisfied by a given square matrix. For instance, it is easily verified that the matrix A in the last example satisfies not only the Cayley-Hamilton equation

$$A^3 - 8A^2 + 13A - 6I = O$$

but also the simpler, quadratic equation

$$A^2 - 7A + 6I = O$$

DEFINITION 1

If A is a square matrix, any polynomial p with the property that $p(A) = O$ is said to annihilate A.

Let us now consider the set of polynomials of minimum degree which **annihilate** a given square matrix A, and, for definiteness, let us assume that by multiplying them by suitable constants their leading coefficients have been made equal to 1. Clearly, all these polynomials are identical In fact, if this is not the case and if there are two such polynomials f and g, then $h = f - g$ is a polynomial whose degree is lower than the degree of f and g such that

$$h(A) = f(A) - g(A) = O - O = O$$

But, by hypothesis, f and g are polynomial annihilators of A of *minimum* degree. Hence we have a contradiction unless h is identically zero, that is, unless f and g are the same.

We are thus justified in introducing the following definition:

DEFINITION 2

The unique polynomial with leading coefficient 1 and of minimum degree which annihilates a square matrix A is called the minimum polynomial of A.

Among the properties of **minimum polynomials**, the following are worthy of mention here

THEOREM 3

Similar matrices have the same minimum polynomial

PROOF Let A and B be similar matrices, so that $B = S^{-1}AS$ Then by Lemma 2, Sec. 11 4, for any polynomial p,

$$p(B) = S^{-1}p(A)S$$

From this we conclude that any polynomial which annihilates A also annihilates B, and conversely Hence, the minimum polynomials of A and B must be the same, as asserted

THEOREM 4

The minimum polynomial of any square matrix A is a divisor of any polynomial which annihilates A.

PROOF Let the minimum polynomial of a matrix A be $f(x)$, and let $\phi(x)$ be any polynomial with the property that $\phi(A) = O$ Then, by the division algorithm of elementary algebra

$$\phi(x) = q(x)f(x) + r(x)$$

where the remainder polynomial $r(x)$ is either identically zero or of lower degree than the divisor polynomial $f(x)$. Then, by Theorem 1, Sec 11.4,

$$\phi(A) = q(A)f(A) + r(A)$$

However, by hypothesis, $\phi(A) = O$ and $f(A) = O$. Hence, $r(A) = O$, and, therefore, $r(x)$ is identically zero, for, if it were not, then it would be a polynomial which annihilated A and whose degree was less than the degree of the minimum polynomial of A, namely, $f(x)$ But if $r(x) \equiv 0$, then the minimum polynomial is a factor of $\phi(x)$, as asserted

THEOREM 5

If the characteristic roots of a matrix A are all distinct, then the characteristic polynomial and the minimum polynomial of A are the same, except possibly for sign.

PROOF Let A be a matrix with distinct characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_n$, and let the characteristic polynomial of A be

$$f(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda)$$

Then, since the minimum polynomial of A, say $g(\lambda)$, must be a factor of $f(\lambda)$, it follows that, if $f(\lambda)$ and $g(\lambda)$ differ in more than sign, then $g(\lambda)$ must be the product of some but not all of the factors of $f(\lambda)$ Specifically, suppose that $g(\lambda)$ does not contain the factor $(\lambda_i - \lambda)$ Now, by Theorem 5, Sec 11 4, the characteristic roots of the matrix $g(A)$ are $g(\lambda_1), g(\lambda_2), \ldots, g(\lambda_n)$ However, since $g(\lambda)$ does not contain the factor $(\lambda_i - \lambda)$, it follows that $g(\lambda_i) \neq 0$ Hence, $g(A)$ has at least one nonzero characteristic root. But, if this is the case, then $g(A)$ is not a null matrix; that is, $g(A) \neq O$, contrary to the hypothesis that g is the minimum polynomial of A. This contradiction shows that $f(\lambda)$ and $g(\lambda)$ cannot differ except possibly in sign, and the theorem is established

As an interesting application of the theory of the minimum polynomial of a matrix, we have the following result

THEOREM 6

If A is a square matrix and if $f(x)$ and $g(x)$ are scalar polynomials such that $g(A)$ is nonsingular, then $f(A)/g(A)$ is equal to a polynomial in A

PROOF Since, by definition, $f(A)/g(A) = f(A)g^{-1}(A)$, it is clearly sufficient to prove that $g^{-1}(A)$ is a polynomial in A To do this, let

$$\phi(x) = x^k + c_1x^{k-1} + \cdots + c_{k-1}x + c_k$$

be the minimum polynomial of the matrix $G = g(A)$ Then

$$\phi(G) = G^k + c_1G^{k-1} + \cdots + c_{k-1}G + c_kI = O$$

and from this, by multiplying through by G^{-1} and transposing, we obtain

$$c_kG^{-1} = -(G^{k-1} + c_1G^{k-2} + \cdots + c_{k-1}I)$$

Now $c_k \neq 0$, for otherwise the right-hand side of the last equation is a polynomial which annihilates G and whose degree is less than the degree k of the minimum polynomial of G Hence, we can divide by c_k and obtain G^{-1} as a polynomial in A Finally, substituting $g(A)$ for G in the expression for G^{-1}, we obtain $G^{-1} = g^{-1}(A)$ as a polynomial in A, as required It is important to note that the structure of the polynomial in A to which $g^{-1}(A)$ is equal depends upon A as well as upon g Hence, if $f(A)/g(A) = h(A)$, we cannot conclude that for another matrix B we necessarily have $f(B)/g(B) = h(B)$

As we have seen, by successive applications of the Cayley-Hamilton theorem it is possible to reduce any polynomial in an (n,n) matrix to another polynomial in A whose degree is at most

$n - 1$. The use of the Cayley-Hamilton theorem is not always the most convenient way to accomplish this reduction, however, and, when the characteristic values of A are all distinct, it is sometimes easier to proceed as follows:

Knowing that, for any polynomial p and any (n,n) matrix A, the matrix $p(A)$ can be expressed as a polynomial, say $\phi(A)$, of degree at most $n - 1$, let us write

$$\begin{aligned}\phi(\lambda) = c_1[(\lambda - \lambda_2)(\lambda - \lambda_3) \cdots (\lambda - \lambda_n)] \\ + c_2[(\lambda - \lambda_1)(\lambda - \lambda_3) \cdots (\lambda - \lambda_n)] + \cdots \\ + c_n[(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_{n-1})]\end{aligned}$$

where the ith term on the right is the product of all the factors of the characteristic polynomial of A except $\lambda - \lambda_i$. Clearly, if $c_1, c_2, \ldots, c_n$ are arbitrary and the λ's are all distinct, the right-hand side is an arbitrary polynomial of degree $n - 1$. Then,

(1) $$\begin{aligned}p(A) = c_1[(A - \lambda_2 I)(A - \lambda_3 I) \cdots (A - \lambda_n I)] \\ + c_2[(A - \lambda_1 I)(A - \lambda_3 I) \cdots (A - \lambda_n I)] + \cdots \\ + c_n[(A - \lambda_1 I)(A - \lambda_2 I) \cdots (A - \lambda_{n-1} I)]\end{aligned}$$

Now, if X_k is a characteristic vector of A corresponding to the characteristic value λ_k, it follows that

(2) $$(A - \lambda_k I)X_k = O$$

Moreover,

$$\begin{aligned}(A - \lambda_j I)X_k &= [A - \lambda_k I + (\lambda_k - \lambda_j)I]X_k \\ &= (A - \lambda_k I)X_k + (\lambda_k - \lambda_j)X_k\end{aligned}$$

(3) $$= (\lambda_k - \lambda_j)X_k$$

Hence, if we postmultiply Eq. (1) by X_k and simplify the products by successively applying Eqs. (2) and (3), we find that every product vanishes except the kth, and we have

(4) $$p(A)X_k = c_k[(\lambda_k - \lambda_1) \cdots (\lambda_k - \lambda_{k-1})(\lambda_k - \lambda_{k+1}) \cdots (\lambda_k - \lambda_n)]X_k$$

Furthermore, according to Corollary 2, Theorem 5, Sec. 11.4, X_k is a characteristic vector of the matrix $p(A)$ corresponding to the characteristic value $p(\lambda_k)$. Therefore,

$$[p(A) - p(\lambda_k)I]X_k = O \qquad \text{or} \qquad p(A)X_k = p(\lambda_k)X_k$$

Thus, Eq. (4) becomes

$$p(\lambda_k)X_k = c_k[(\lambda_k - \lambda_1) \cdots (\lambda_k - \lambda_{k-1})(\lambda_k - \lambda_{k+1}) \cdots (\lambda_k - \lambda_n)]X_k$$

which implies that

$$p(\lambda_k) = c_k \prod_{\substack{i=1 \\ i \neq k}}^{n} (\lambda_k - \lambda_i) \qquad \text{or} \qquad c_k = \frac{p(\lambda_k)}{\displaystyle\prod_{\substack{i=1 \\ i \neq k}}^{n} (\lambda_k - \lambda_i)}$$

Therefore, substituting these values for the c's into Eq. (1), we obtain the identity

$$(5) \qquad p(A) = \sum_{k=1}^{n} \left[\frac{p(\lambda_k)}{\prod\limits_{\substack{i=1\\ i\neq k}}^{n} (\lambda_k - \lambda_i)} \prod_{\substack{i=1\\ i\neq k}}^{n} (A - \lambda_i I) \right]$$

This important result is known as **Sylvester's identity.**[*] It may be extended to cover the case in which the characteristic values of A are not all distinct, but we shall not undertake this extension [†]

EXAMPLE 2

If

$$A = \left\| \begin{matrix} -15 & -14 & -40 \\ 6 & 7 & 14 \\ 5 & 4 & 14 \end{matrix} \right\|$$

express $p(A) = A^6 - 6A^5 + 12A^4 - 12A^3 + 12A^2 - 8A + 3I$ in as simple a form as possible

By a straightforward calculation we find, for the characteristic equation of A,

$$|A - \lambda I| = \begin{vmatrix} -15-\lambda & -14 & -40 \\ 6 & 7-\lambda & 14 \\ 5 & 4 & 14-\lambda \end{vmatrix} = -\lambda^3 + 6\lambda^2 - 11\lambda + 6 = 0$$

Hence, the characteristic values of A are $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$ Therefore,

$$p(\lambda_1) = p(1) = 2$$
$$p(\lambda_2) = p(2) = 3$$
$$p(\lambda_3) = p(3) = 6$$

and, substituting into Eq (5),

$$p(A) = \frac{2}{(1-2)(1-3)}(A - 2I)(A - 3I) + \frac{3}{(2-1)(2-3)}(A - I)(A - 3I)$$
$$+ \frac{6}{(3-1)(3-2)}(A - I)(A - 2I) = A^2 - 2A + 3I$$

EXERCISES

1 Without using the Cayley-Hamilton theorem, prove that every (n,n) matrix satisfies a polynomial equation of degree at most n^2

2 Using the Cayley-Hamilton theorem, find the inverse of each of the following matrices

a $\left\| \begin{matrix} 2 & -4 & -4 \\ 1 & -4 & -5 \\ -1 & 4 & 5 \end{matrix} \right\|$ **b** $\left\| \begin{matrix} 2 & 1 & 1 \\ 1 & 4 & 3 \\ -1 & -1 & 0 \end{matrix} \right\|$ **c** $\left\| \begin{matrix} -4 & -9 & -3 \\ 1 & 4 & 1 \\ 3 & 3 & 2 \end{matrix} \right\|$

3 Find the minimum polynomial of each of the following matrices

a $\left\| \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} \right\|$ **b** $\left\| \begin{matrix} -1 & 2 & 2 \\ 2 & -1 & -2 \\ -2 & 2 & 3 \end{matrix} \right\|$

c $\left\| \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 3 \end{matrix} \right\|$ **d** $\left\| \begin{matrix} 7 & 2 & -2 \\ -6 & -1 & 2 \\ 6 & 2 & -1 \end{matrix} \right\|$

---

[*] Named for the English algebraist J J Sylvester (1814–1897)

[†] See, for instance, W J Duncan, R A Fraser, and A R Collar, "Elementary Matrices," pp 78–79, Cambridge University Press, New York, 1938.

4 Using both the Cayley-Hamilton theorem and Sylvester's identity, evaluate

$$p(A) = A^5 - A^4 - 2A^3 + A^2 + A - 3I$$

for each of the following matrices A ·

a $\begin{Vmatrix} 3 & -2 \\ 5 & -4 \end{Vmatrix}$ **b** $\begin{Vmatrix} 4 & -1 \\ 6 & -1 \end{Vmatrix}$ **c** $\begin{Vmatrix} -2 & -6 \\ 2 & 5 \end{Vmatrix}$

d $\begin{Vmatrix} 2 & -4 & -4 \\ 1 & -4 & -5 \\ -1 & 4 & 5 \end{Vmatrix}$ **e** $\begin{Vmatrix} -1 & 1 & 1 \\ 2 & -2 & -3 \\ -2 & 2 & 3 \end{Vmatrix}$

5 If $f(x) = x/(x+4)$, express $f(A)$ as a polynomial in A, for each of the following matrices A.

a $\begin{Vmatrix} 4 & 1 \\ -3 & 0 \end{Vmatrix}$ **b** $\begin{Vmatrix} -1 & -2 \\ 3 & 4 \end{Vmatrix}$ **c** $\begin{Vmatrix} 4 & 6 \\ -3 & -5 \end{Vmatrix}$

d $\begin{Vmatrix} -4 & -9 & -3 \\ 1 & 4 & 1 \\ 3 & 3 & 2 \end{Vmatrix}$ **e** $\begin{Vmatrix} 2 & 1 & 1 \\ 1 & 4 & 3 \\ -1 & -1 & 0 \end{Vmatrix}$

11.6

Infinite series of matrices

In the last two sections we have considered polynomial and rational fractional functions of a square matrix In this section we shall conclude our survey of the theory of matrices by investigating briefly, and for the most part without proof, infinite series of matrices Once we have suitable criteria for the convergence of infinite series of matrices, it will then be possible to define and study transcendental functions of square matrices, such as e^A, sin A, and cos A, by allowing the corresponding scalar series to have matric arguments As our experience with scalar series suggests, we must begin with the concept of the **convergence of a sequence of matrices**:

DEFINITION 1

If $A_1, A_2, \quad ,A_n, \quad$ is a sequence of (p,q) matrices, if $(a_{ij})_n$ is the element in the ith row and jth column of A_n, and if a_{ij} is the corresponding element in a (p,q) matrix A, the sequence $A_1, A_2, \quad , A_n, \quad$ is said to converge to the matrix A if and only if, for all values of i and j, $\lim_{n\to\infty} (a_{ij})_n = a_{ij}$

We indicate that a sequence of matrices $\{A_n\}$ converges to a matrix A by writing $A_n \to A$ A sequence of matrices which does not converge is said to **diverge** or to be **divergent.**

According to Definition 1, the convergence of a sequence of (p,q) matrices depends on the convergence of pq scalar sequences. Hence, it is only an easy application of familiar ideas to prove the following results

LEMMA 1

If $\{A_n\}$ and $\{B_n\}$ are two sequences of (p,q) matrices and if $A_n \to A$ and $B_n \to B$, then, for all scalar constants α and β, $\alpha A_n + \beta B_n \to \alpha A + \beta B$.

LEMMA 2

If $\{A_n\}$ and $\{B_n\}$ are two sequences of suitably conformable matrices and if $A_n \to A$ and $B_n \to B$, then $A_nB_n \to AB$.

LEMMA 3

If $\{A_n\}$ is a sequence of (p,q) matrices and if $A_n \to A$, then, for suitably conformable matrices R and S, $RA_nS \to RAS$.

PROOF From the definition of matric multiplication, the element in the ith row and jth column of the product RA_nS is

$$\sum_{k=1}^{p} \sum_{l=1}^{q} r_{ik}(a_{kl})_n s_{lj}$$

Since r_{ik} and s_{lj} are clearly independent of n and since, by hypothesis, $(a_{kl})_n \to a_{kl}$, it follows that this finite sum converges to

$$\sum_{k=1}^{p} \sum_{l=1}^{q} r_{ik}a_{kl}s_{lj}$$

which is the element in the ith row and jth column of the product RAS, as asserted

Using the fact that any square matrix is similar to an upper triangular matrix, it is possible to prove the following important theorem.*

THEOREM 1

If A is a square matrix, a necessary and sufficient condition that $A^n \to 0$ is that the absolute value of each characteristic root of A be less than 1

We are now in a position to define the **convergence of an infinite series of matrices**:

DEFINITION 2

The series of matrices $\sum_{m=0}^{\infty} c_mA_m$ is said to converge to the sum S if and only if the sequence of partial sums

$$\{S_n\} = \left\{ \sum_{m=0}^{n} c_mA_m \right\}$$

converges to S as n becomes infinite

DEFINITION 3

A series of matrices $\sum_{m=0}^{\infty} c_mA_m$ is said to converge absolutely if and only if each scalar series $\sum_{m=0}^{\infty} c_m(a_{ij})_m$ is absolutely convergent.

As criterion for the absolute convergence of a series of matrices, we have the result contained in the following theorem

---

* See, for instance, Mirsky, "Linear Algebra," p. 328.

THEOREM 2

If a_m is the element of maximum absolute value in the matrix A_m, then a necessary and sufficient condition that the series $\sum_{m=0}^{\infty} c_m A_m$ converge absolutely is that the scalar series $\sum_{m=0}^{\infty} |c_m|\ |a_m|$ converge

PROOF To prove the necessity of the condition of the theorem, let us assume that the given series $\sum_{m=0}^{\infty} c_m A_m$ is absolutely convergent Then, if A_m is a (p,q) matrix, it follows that, for all values of i and j such that $1 \leqq i \leqq p$ and $1 \leqq j \leqq q$, the series $\sum_{m=0}^{\infty} c_m(a_{ij})_m$ is absolutely convergent, that is, $\sum_{m=0}^{\infty} |c_m| \cdot |(a_{ij})_m|$ converges. Let L ($\geqq 0$) be the largest of the sums to which these pq series converge. Then, clearly, for all values of i and j and for every $n > 0$,

$$\sum_{m=0}^{n} |c_m|\ |(a_{ij})_m| \leqq L$$

If we now sum the last inequality for all values of i and j, we obtain

$$\sum_{i,j} \sum_{m=0}^{n} |c_m| \cdot |(a_{ij})_m| \leqq \sum_{i,j} L = pqL$$

From this, by reversing the order of summation on the left and noting that $|c_m|$ is independent of i and j, we have

$$\sum_{m=0}^{n} |c_m| \sum_{i,j} |(a_{ij})_m| \leqq pqL \tag{1}$$

Now the absolute value of the element of largest absolute value in A_m surely cannot exceed the sum of the absolute values of *all* the elements in A_m; that is,

$$|a_m| \leqq \sum_{i,j} |(a_{ij})_m|$$

Hence, using this to underestimate the left member of (1), we have

$$\sum_{m=0}^{n} |c_m|\ |a_m| \leqq pqL$$

Thus the series $\sum_{m=0}^{\infty} |c_m| \cdot |a_m|$ converges, since its partial sums form a bounded monotonically increasing sequence, and the necessity of the condition of the theorem is established.

To prove the sufficiency of the condition of the theorem, let us suppose that $\sum_{m=0}^{\infty} |c_m| \cdot |a_m|$ converges. Then, since $|(a_{ij})_m| \leqq |a_m|$ for all i and j, it is clear that $|c_m|\ |(a_{ij})_m| \leqq |c_m|\ |a_m|$ for all i and j. Hence, by an easy application of the comparison test for the convergence of scalar series, it follows that $\sum_{m=0}^{\infty} |c_m| \cdot |(a_{ij})_m|$ converges for all i and j, which proves that the given series of matrices converges absolutely, as asserted.

If a series of matrices is absolutely convergent, it follows at once from the corresponding properties of scalar series that the matric terms can be rearranged at pleasure without in any way affecting the sum of the series

With the exception of Theorem 1, all the observations we have so far made about sequences and series of matrices apply to matrices of any shape However in most applications we are concerned only with series of square matrices and, in particular, with power series of such matrices The fundamental theorem on matric power series is the following:*

THEOREM 3

If the absolute value of each characteristic root of a square matrix A is less than the radius of convergence of the scalar power series $\phi(z) = \sum_{m=0}^{\infty} c_m z^m$, then the matric power series $\phi(A) = \sum_{m=0}^{\infty} c_m A^m$ converges If the absolute value of at least one characteristic root of A is greater than the radius of convergence of $\phi(z)$, then $\phi(A)$ diverges In particular cases, if the absolute value of the characteristic root of largest absolute value is equal to the radius of convergence of $\phi(z)$, the matric series $\phi(A)$ may either converge or diverge

COROLLARY 1

If $\phi(z)$ converges for all values of z, then $\phi(A)$ converges for all square matrices A

COROLLARY 2

If $\phi(A) = \sum_{m=0}^{\infty} c_m A^m$ converges and if A is similar to a diagonal matrix, that is, if

$$S^{-1}AS = D = \left\| \begin{array}{ccc} \lambda_1 & & \bigcirc \\ & \lambda_2 & \\ \bigcirc & & \lambda_n \end{array} \right\|$$

then

$$\phi(A) = S \left\| \begin{array}{ccc} \phi(\lambda_1) & & \bigcirc \\ & \phi(\lambda_2) & \\ \bigcirc & & \phi(\lambda_n) \end{array} \right\| S^{-1}$$

PROOF Clearly, the partial sums of the series $\phi(A)$ are all polynomials in A Hence, by Theorem 6, Sec 11 4,

$$\phi_N(A) = \sum_{m=1}^{N} c_m A^m = S \left\| \begin{array}{ccc} \phi_N(\lambda_1) & & \bigcirc \\ & \phi_N(\lambda_2) & \\ \bigcirc & & \phi_N(\lambda_n) \end{array} \right\| S^{-1}$$

---

* See, for instance, Mirsky, "Linear Algebra," p. 332.

Now, by hypothesis, $\phi(A)$ converges. Hence, each of the scalar series $\phi(\lambda_i)$ must converge; that is, for each i, $\phi_N(\lambda_i) \to \phi(\lambda_i)$. Therefore, as $N \to \infty$,

$$\begin{Vmatrix} \phi_N(\lambda_1) & & \bigcirc \\ & \phi_N(\lambda_2) & \\ \bigcirc & & \phi_N(\lambda_n) \end{Vmatrix} \to \begin{Vmatrix} \phi(\lambda_1) & & \bigcirc \\ & \phi(\lambda_2) & \\ \bigcirc & & \phi(\lambda_n) \end{Vmatrix}$$

and we have, by Lemma 3,

$$\phi(A) = S \begin{Vmatrix} \phi(\lambda_1) & & \bigcirc \\ & \phi(\lambda_2) & \\ \bigcirc & & \phi(\lambda_n) \end{Vmatrix} S^{-1} \qquad \text{as asserted.}$$

The last result provides us with a useful method of evaluating the sum of certain matric power series, which is sometimes preferable to the use of the Cayley-Hamilton theorem or Sylvester's identity

EXAMPLE 1

What is e^A if $A = \begin{Vmatrix} 0 & -2 \\ 1 & 3 \end{Vmatrix}$?

The given matrix A is the one we considered in Example 3, Sec 11 4 Hence from that example we know that the characteristic equation of A is $\lambda^2 - 3\lambda + 2 = 0$ and that $\lambda_1 = 1$, $\lambda_2 = 2$, and

$$S^{-1}AS = \begin{Vmatrix} 1 & 0 \\ 0 & 2 \end{Vmatrix} \qquad \text{where } S = \begin{Vmatrix} 2 & 1 \\ -1 & -1 \end{Vmatrix} \qquad \text{and} \qquad S^{-1} = \begin{Vmatrix} 1 & 1 \\ -1 & -2 \end{Vmatrix}$$

Therefore, by the last corollary,

$$e^A = S \begin{Vmatrix} \phi(\lambda_1) & 0 \\ 0 & \phi(\lambda_2) \end{Vmatrix} S^{-1} = \begin{Vmatrix} 2 & 1 \\ -1 & -1 \end{Vmatrix} \begin{Vmatrix} e & 0 \\ 0 & e^2 \end{Vmatrix} \begin{Vmatrix} 1 & 1 \\ -1 & -2 \end{Vmatrix}$$

$$= \begin{Vmatrix} 2e - e^2 & 2e - 2e^2 \\ -e + e^2 & -e + 2e^2 \end{Vmatrix}$$

To evaluate e^A by means of the Cayley-Hamilton theorem, we must simplify each of the powers of A in the expansion of e^A, using the relation $A^2 - 3A + 2I = O$, or

$$A^2 = 3A - 2I$$

At first glance this would seem to be a very tedious process, since arbitrarily high powers of A are involved However, we may shorten the work appreciably by proceeding inductively From the fact that the characteristic equation of A is quadratic, we know that any positive integral power of A can be expressed as a linear binomial in A Hence, if we assume

$$A^n = a_nA + b_nI \qquad n = 2, 3, 4,$$

we have

$$A^{n+1} = a_{n+1}A + b_{n+1}I = A\ A^n = a_nA^2 + b_nA = a_n(3A - 2I) + b_nA$$
$$= (3a_n + b_n)A - 2a_nI$$

Therefore the a's and b's satisfy the recurrence relations*

$$a_{n+1} = 3a_n + b_n \qquad \text{and} \qquad b_{n+1} = -2a_n$$

---

* These are examples of what are known as *difference equations*, see Sec 4.5

with, of course, the initial conditions $a_2 = 3$, $b_2 = -2$. From these it is easy to verify that

$$\begin{array}{ll} a_2 = 3 & b_2 = -2 \\ a_3 = 7 & b_3 = -6 \\ a_4 = 15 & b_4 = -14 \\ a_5 = 31 & b_5 = -30 \\ & \cdot \\ a_n = 2^n - 1 & b_n = -2^n + 2 \end{array}$$

Hence,

$$\begin{aligned} e^A &= I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \\ &= I + A + \frac{(2^2 - 1)A + (-2^2 + 2)I}{2!} + \frac{(2^3 - 1)A + (-2^3 + 2)I}{3!} + \\ &= A\left(1 + \frac{2^2 - 1}{2!} + \frac{2^3 - 1}{3!} + \qquad\right) + I\left(1 + \frac{2 - 2^2}{2!} + \frac{2 - 2^3}{3!} + \qquad\right) \\ &= A\left[\left(1 + \frac{2}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \qquad\right) - \left(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \qquad\right)\right] \\ &\quad + I\left[2\left(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \qquad\right) - \left(1 + \frac{2}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \qquad\right)\right] \\ &= A(e^2 - 2) - I(e^2 - 2e) \\ &= (e^2 - e)\begin{Vmatrix} 0 & -2 \\ 1 & 3 \end{Vmatrix} - (e^2 - 2e)\begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix} \\ &= \begin{Vmatrix} 2e - e^2 & 2e - 2e^2 \\ -e + e^2 & -e + 2e^2 \end{Vmatrix} \end{aligned}$$

as before Finally, using Sylvester's formula, we have

$$\begin{aligned} \phi(\ 1\ ) &= \frac{\phi(\lambda_1)}{\lambda_1 - \lambda_2}(A - \lambda_2 I) + \frac{\phi(\lambda_2)}{\lambda_2 - \lambda_1}(A - \lambda_1 I) \\ &= \frac{e}{-1}(A - 2I) + \frac{e^2}{1}(A - I) \\ &= A(e^2 - e) - I(e^2 - 2e) \end{aligned}$$

which is the first expression we obtained using the Cayley-Hamilton theorem

The question of whether or not scalar identities such as $e^x e^y = e^{x+y}$, $\sin 2x = 2 \sin x \cos x$, and

$$\cos (x + y) = \cos x \cos y - \sin x \sin y$$

remain valid when x and y are replaced by square matrices is obviously an important one. We cannot investigate the matter here, but the applications one is likely to encounter are covered by the following theorem *

THEOREM 4

For the elementary transcendental functions, scalar identities in a single variable remain valid when the scalar variable is replaced by a square matrix Scalar identities in two variables remain valid when the scalar variables are replaced by square matrices only when the two matrices commute.

---

* See, for example, Mirsky, "Linear Algebra," pp. 338–341

EXERCISES

1 Prove Lemma 1

2 Prove Lemma 2.

3 For the matrix A of Example 1, compute e^{-A} and verify that $e^{A}e^{-A} = I$

4 Using Sylvester's identity, compute e^{A}, $\cos A$, and $\sin A$ for each of the following matrices A:

a $\begin{Vmatrix} 3 & -1 \\ 6 & -2 \end{Vmatrix}$ **b** $\begin{Vmatrix} 2 & -1 \\ 4 & -3 \end{Vmatrix}$

c $\begin{Vmatrix} 2 & 1 & 1 \\ 1 & 4 & 3 \\ -1 & -1 & 0 \end{Vmatrix}$ **d** $\begin{Vmatrix} -1 & 1 & 1 \\ 2 & -2 & -3 \\ -2 & 2 & 3 \end{Vmatrix}$

5 Using both the Cayley-Hamilton theorem and Sylvester's identity, evaluate e^{A} for $A = \begin{Vmatrix} 1 & -1 \\ -1 & 0 \end{Vmatrix}$.

6 For each of the following matrices A, verify that $\sin 2A = 2 \sin A \cos A$.

a $\begin{Vmatrix} 3 & 2 \\ -3 & -2 \end{Vmatrix}$ **b** $\begin{Vmatrix} 2 & 1 \\ -3 & -2 \end{Vmatrix}$

c $\begin{Vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{Vmatrix}$ **d** $\begin{Vmatrix} -1 & 2 & 2 \\ 2 & -1 & -2 \\ -2 & 2 & 3 \end{Vmatrix}$

7 If $A = \begin{Vmatrix} 2 & 1 \\ -2 & -1 \end{Vmatrix}$ and $B = \begin{Vmatrix} 1 & 0 \\ -2 & -1 \end{Vmatrix}$, verify that none of the following relations holds

$$\sin (A \pm B) = \sin A \cos B \pm \cos A \sin B$$

$$\cos (A \pm B) = \cos A \cos B \mp \sin A \sin B$$

8 If $A = \begin{Vmatrix} 2/3 & -1/3 \\ -2/3 & 1/3 \end{Vmatrix}$ and $B = \begin{Vmatrix} 1/3 & -2/3 \\ -1/3 & -1/3 \end{Vmatrix}$, verify that each of the following relations holds

$$\sin (A \pm B) = \sin A \cos B \pm \cos A \sin B$$

$$\cos (A \pm B) = \cos A \cos B \mp \sin A \sin B$$

9 Determine for what values of x and y, if any, the matrices

$$A = \begin{Vmatrix} x & x - 1 \\ -x & 1 - x \end{Vmatrix} \qquad \text{and} \qquad B = \begin{Vmatrix} y & y - 1 \\ -1 - y & -y \end{Vmatrix}$$

commute, and verify that, under these conditions,

$$\sin (A \pm B) = \sin A \cos B \pm \cos A \sin B$$

$$\cos (A \pm B) = \cos A \cos B \mp \sin A \sin B$$

10 If $A = \begin{Vmatrix} 1 & 0 & 2 \\ 0 & -1 & -2 \\ 2 & -2 & 0 \end{Vmatrix}$, show that $\sin A = A \dfrac{\sin 3}{3}$ Obtain a similar expression for $\cos A$, and verify that $\cos^2 A + \sin^2 A = I$

CHAPTER TWELVE

Vector Analysis

12.1

The algebra of vectors

In Sec. 10.2, in our discussion of determinants and matrices, we introduced the concept of a vector as an ordered set of n quantities, say $(a_1,a_2, \ldots ,a_n)$. In the present chapter we shall undertake the study of what is known as *vector analysis*, using the more traditional (and limited) definition of a vector as a quantity, such as force, velocity, or acceleration, which possesses both magnitude and direction. Although this approach has been all but abandoned in pure mathematics because it is unnecessarily restricted, it is still the usual approach in physics and in engineering, and the results to which it leads are of great utility in these fields.

Almost any physical discussion will involve, in addition to vector quantities, other quantities, such as volume, mass, and work, which possess only magnitude and are known as **scalars.** To distinguish vectors from scalars we shall consistently write the former in boldface type, thus, $\mathbf{V}$. This is a rather common notation, although some authors indicate that a symbol stands for a vector quantity by putting an arrow above the symbol, thus, $\vec{V}$.

A scalar quantity can be adequately represented by a mark on a fixed scale. To represent a vector quantity, however, we must use a directed line segment whose direction is the same as the direction of the vector and whose length is equal (on some convenient scale) to the magnitude of the vector. For convenience, we shall often refer to the representative line segment as though it were the vector itself. The magnitude or length of a vector $\mathbf{A}$ is called the **absolute value** of the vector and is indicated either by enclosing the symbol for the vector between ordinary absolute-value bars or simply by setting the symbol for the vector in ordinary rather than boldface type. Thus,

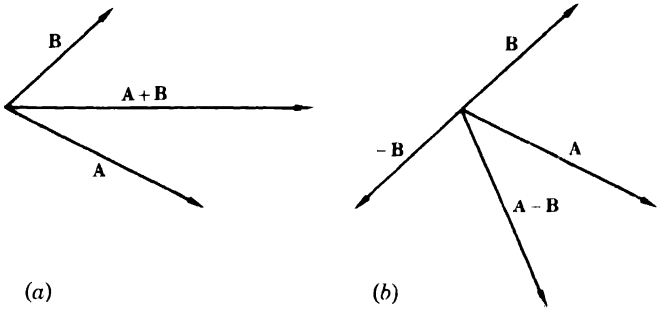
$$A = |\mathbf{A}|$$

represents the magnitude, or absolute value, of the vector **A**. Regardless of its direction, a vector whose length, or absolute value, is unity is called a **unit vector**. A vector is said to be **zero** if and only if its absolute value is zero The direction of a zero vector is undefined.

Two vectors whose magnitudes, or lengths, are equal and whose directions are the same are said to be **equal**, regardless of the points in space from which they may be drawn * If two vectors have the same length but are oppositely directed, either is said to be the **negative** of the other.

The **sum** of two vectors **A** and **B** is defined by the familiar parallelogram law, i.e , if **A** and **B** are drawn from the same point, or origin, and if the parallelogram having **A** and **B** as adjacent sides is constructed, then the sum $\mathbf{A} + \mathbf{B}$ is the vector represented by the diagonal of this parallelogram which passes through the common origin of **A** and **B** (Fig 12 1*a*) From this definition it is evident that

FIGURE 12.1
The addition and subtraction of vectors



(*a*) (*b*)

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

i e , that *vector addition is commutative*, and that

$$\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$$

i e , that *vector addition is associative* By the **difference** of two vectors **A** and **B**, we mean the sum of the first and the negative of the second, i e ,

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B})$$

(Fig 12 1*b*) By the **product of a scalar** a **and a vector A** we mean the vector $a\mathbf{A}$ whose length is equal to the product of $|a|$ and the magnitude of **A** and whose direction is the same as the direction of **A** if a is positive and opposite to it if a is negative.

---

* In other words, a vector quantity can be represented equally well by any of infinitely many equivalent line segments, all having the same length and the same direction It is, therefore, customary to say that a vector can be moved parallel to itself without change In some applications, however, as for instance in dealing with forces whose points of application or lines of action cannot be shifted, it is necessary to think of a vector as fixed, or at least limited in position Such vectors are usually said to be **bound**, in contrast to unrestricted vectors, which are said to be **free**.

In addition to the product of a scalar and a vector, two other types of product are defined in vector analysis. The first of these is the **scalar** or **dot** or **inner product,** indicated by placing a dot between the two factors. By definition, this is a *scalar* equal to the product of the absolute values of the two vector factors and the cosine of the angle between their positive directions; i.e.,

(1) $$\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}|\,|\mathbf{B}|\cos\theta = AB\cos\theta$$

Since $|\mathbf{A}|\cos\theta$ is just the projection of the vector **A** in the direction of **B** and since $|\mathbf{B}|\cos\theta$ is the projection of the vector **B** in the direction of **A**, it follows that *the dot product of two vectors is equal to the length of either of them multiplied by the projection of the other upon it* (Fig. 12.2). Two particular cases of this are worthy of note: If one of the vectors, say **A**, is of unit length, then $\mathbf{A}\cdot\mathbf{B}$ becomes simply

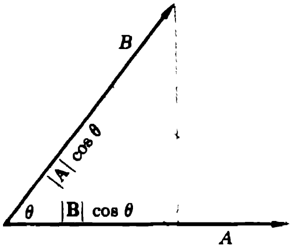


FIGURE 12.2 The geometrical interpretation of the scalar product

$$|\mathbf{B}|\cos\theta = B\cos\theta$$

which is just the projection, or **component,** of **B** in the direction of the unit vector **A**. On the other hand, if $\mathbf{A} = \mathbf{B}$, then, obviously, $\cos\theta = \cos 0 = 1$, and we have

(2) $$\mathbf{A}\cdot\mathbf{A} = |\mathbf{A}|^2 = A^2$$

From the relation between dot products and projections it is easy to show that *dot multiplication is distributive over addition,* i.e.,

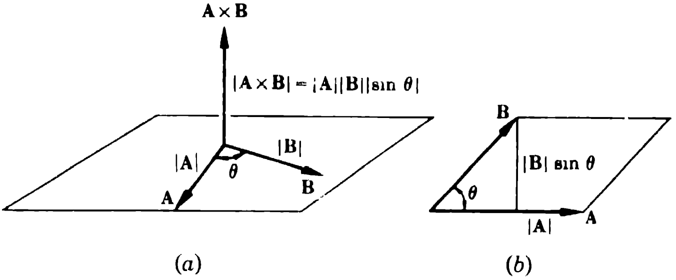
(3) $$\mathbf{A}\cdot(\mathbf{B}+\mathbf{C}) = \mathbf{A}\cdot\mathbf{B} + \mathbf{A}\cdot\mathbf{C}$$

Moreover, from the definitive relation (1) it is clear that *dot multiplication is commutative,* i.e.,

(4) $$\mathbf{A}\cdot\mathbf{B} = \mathbf{B}\cdot\mathbf{A}$$

However, if the dot product of two vectors is zero, it does not follow that one or the other of the factors is zero, for there is a third possibility, namely, $\cos\theta = 0$. Thus *if* $\mathbf{A}\cdot\mathbf{B} = 0$, *then either at least one of the vectors* (**A**,**B**) *is zero or* **A** *and* **B** *are perpendicular*

The third type of product with which we shall deal is the **vector,** or **cross product,** indicated by placing a cross between the factors.* If **A** and **B** are the factors, then by definition $\mathbf{A} \times \mathbf{B}$ is a *vector* **V** whose absolute value is the product of the absolute values of **A**, **B**, and the sine of the angle between them, and whose direction is perpendicular to the plane determined by **A** and **B** and so sensed that a right-handed screw turned from **A** toward **B** through the smaller of the angles between these vectors would advance in the direction of **V** (Fig. 12.3*a*). Since $|\mathbf{B}|\ |\sin\theta|$ is the projection of **B** in a direction perpendicular to **A**, or, in other words, is the altitude of the parallelogram determined when **A** and **B** are drawn from a common point, it follows that the magnitude of $\mathbf{A} \times \mathbf{B}$, namely, $|\mathbf{A}|\ (|\mathbf{B}|\ |\sin\theta|)$, is equal to the area of this parallelogram (Fig 12.3*b*)



(*a*) (*b*)

FIGURE 12.3 The geometrical interpretation of the vector product.

From the relation between cross products and areas it is easy to show that *cross multiplication is distributive over addition*, i.e.,

$$\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C} \tag{5}$$

However, since the direction of $\mathbf{A} \times \mathbf{B}$ is determined by the right-hand rule, it is clear that interchanging **A** and **B** reverses the direction, or sign, of their product Hence, *cross multiplication is not commutative*, and we have, in fact,

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A} \tag{6}$$

Multiplication in which products obey this rule is sometimes said to be **anticommutative.**

From the foregoing it is clear that we must be careful to preserve the proper order of factors in any expression involving vector multiplication Moreover, if $\mathbf{A} \times \mathbf{B} = 0$, we cannot conclude that either **A** or **B** is zero, for this product will also vanish if $\sin\theta = 0$ Hence *if* $\mathbf{A} \times \mathbf{B} = 0$, *then either at least one of the vectors* (**A**,**B**) *is zero or* **A** *and* **B** *are parallel.*

It is often convenient to be able to refer vector expressions to a cartesian frame of reference To provide for this we define **i**, **j**, and **k** to be vectors of unit length directed, respectively, along the positive x-, y-, and z-axes of a right-handed coordinate system. Then $x\mathbf{i}$, $y\mathbf{j}$, and $z\mathbf{k}$ represent vectors of lengths x, y, and z whose directions are those of the respective axes, and from the definition of vector addition it is evident that the vector joining the origin to a general point $P:(x,y,z)$ (Fig. 12.4) can be written

---

* Meaning has also been given to the symbol **AB**, and in fact under the name **dyad** such combinations have been extensively studied, as, for instance, in Gibbs-Wilson, "Vector Analysis," Yale University Press, New Haven, Conn., 1929 We shall not consider them in our work, however, since they are actually special cases of what are known as *tensors*, which we shall consider from a somewhat different point of view in the next chapter For us, the only possible product-type combinations of two vectors will be the dot and cross products themselves

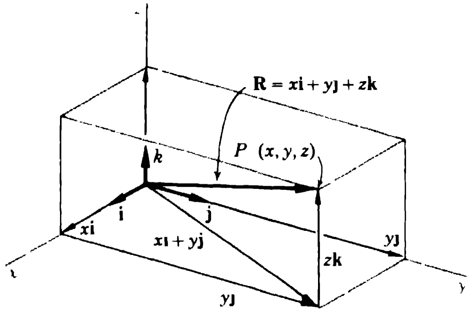


FIGURE 12.4 The representation of a vector as a linear combination of the unit vectors **i**, **j**, **k**

(7) $$\mathbf{R} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

In more general terms, any vector whose components along the axes are, respectively, a_1, a_2, and a_3 can be written

$$\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$$

If, further,

$$\mathbf{B} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$$

then $$\mathbf{A} \pm \mathbf{B} = (a_1 \pm b_1)\mathbf{i} + (a_2 \pm b_2)\mathbf{j} + (a_3 \pm b_3)\mathbf{k}$$

Clearly, *two vectors will be equal if and only if their respective components are equal* Hence, any vector equation implies three scalar equations.

Since the dot product of perpendicular vectors is zero, it follows that

(8) $$\mathbf{i} \cdot \mathbf{j} = \mathbf{j} \cdot \mathbf{k} = \mathbf{k} \cdot \mathbf{i} = 0$$

Moreover, applying (2) to the unit vectors **i**, **j**, **k**, we have

(9) $$\mathbf{i} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = 1$$

Hence, if we write

$$\mathbf{A} \cdot \mathbf{B} = (a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}) \cdot (b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k})$$

and use the fact that dot multiplication is distributive over addition [Eq. (3)] to expand and simplify, we obtain the important result

(10) $$\mathbf{A} \cdot \mathbf{B} = a_1b_1 + a_2b_2 + a_3b_3$$

In particular, taking **B** = **A**, we have

$$\mathbf{A}\cdot\mathbf{A} = |\mathbf{A}|^2 = a_1^2 + a_2^2 + a_3^2$$

or

(11) $$|\mathbf{A}| = \sqrt{a_1^2 + a_2^2 + a_3^2}\dagger$$

On the other hand, if we write $\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}|\,|\mathbf{B}|\cos\theta$ and then solve for $\cos\theta$, using (10) and (11), we obtain the useful formula

(12) $$\cos\theta = \frac{a_1b_1 + a_2b_2 + a_3b_3}{\sqrt{a_1^2 + a_2^2 + a_3^2}\sqrt{b_1^2 + b_2^2 + b_3^2}}$$

a result familiar from analytic gemoetry, where the a's and b's were introduced not as the components of two vectors, but as the direction numbers of two straight lines.

For the cross products of the unit vectors **i**, **j**, **k** we find at once

(13) $$\begin{aligned}\mathbf{i}\times\mathbf{i} &= \mathbf{j}\times\mathbf{j} = \mathbf{k}\times\mathbf{k} = 0\\ \mathbf{i}\times\mathbf{j} &= -\mathbf{j}\times\mathbf{i} = \mathbf{k}\\ \mathbf{j}\times\mathbf{k} &= -\mathbf{k}\times\mathbf{j} = \mathbf{i}\\ \mathbf{k}\times\mathbf{i} &= -\mathbf{i}\times\mathbf{k} = \mathbf{j}\end{aligned}$$

Hence, using (13) and the fact that cross multiplication is distributive over addition, we obtain for

$$\mathbf{A}\times\mathbf{B} \equiv (a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k})\times(b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k})$$

the expression

(14) $$\mathbf{A}\times\mathbf{B} = (a_2b_3 - a_3b_2)\mathbf{i} - (a_1b_3 - a_3b_1)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k}$$

which is precisely the expanded form of the determinant

(15) $$\mathbf{A}\times\mathbf{B} = \begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\end{vmatrix}$$

The anticommutative character of vector multiplication thus corresponds to the fact that interchanging two rows of a determinant changes the sign of the determinant.

EXAMPLE 1

Using vector methods, derive the law of cosines

To do this, let directions be assigned to the sides of the given triangle as in Fig 12 5 Then $\mathbf{C} = \mathbf{A} - \mathbf{B}$, hence,

$$\mathbf{C}\cdot\mathbf{C} = (\mathbf{A} - \mathbf{B})\cdot(\mathbf{A} - \mathbf{B}) = \mathbf{A}\cdot\mathbf{A} - 2\mathbf{A}\cdot\mathbf{B} + \mathbf{B}\cdot\mathbf{B}$$

or, using (1) and (2),

$$C^2 = A^2 + B^2 - 2AB\cos\theta$$

which is the law of cosines

---

† Formula (10) is precisely the scalar product of the two vectors $\mathbf{A} = \|a_1,a_2,a_3\|$ and $\mathbf{B} = \|b_1,b_2,b_3\|$ as we defined it from the more general point of view of Chap 10 (Sec. 10 2) Similarly, Formula (11) gives the length of the vector $\mathbf{A} = \|a_1,a_2,a_3\|$ as we defined it in Chap 10, Sec 10 2.

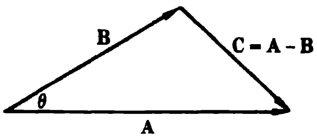


FIGURE 12.5 The triangle used in the vector derivation of the law of cosines

EXAMPLE 2

If (x,y,z) and (x',y',z') are two right-handed coordinate systems having a common origin, obtain by vector methods the transformation equations connecting the two systems of coordinates

To do this, let **i**, **j**, **k** and **i**′, **j**′, **k**′ be unit vectors in the direction of the respective axes (Fig. 12 6), and let P be a general point in space having coordinates (x,y,z) and (x',y',z') in the respective systems Now, the coordinates (x',y',z') are simply the components of the vector OP along the x'-, y'-, z'-axes Hence, if we write

$$\mathbf{R} = OP = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

and observe that the dot products of this vector with the unit vectors **i**′, **j**′, and **k**′ are its components in these directions, we find the required formulas to be

$$\begin{aligned} x' &= \mathbf{R}\cdot\mathbf{i}' = (x\mathbf{i} + y\mathbf{j} + z\mathbf{k})\cdot\mathbf{i}' = x(\mathbf{i}\cdot\mathbf{i}') + y(\mathbf{j}\cdot\mathbf{i}') + z(\mathbf{k}\cdot\mathbf{i}') \\ y' &= \mathbf{R}\cdot\mathbf{j}' = (x\mathbf{i} + y\mathbf{j} + z\mathbf{k})\cdot\mathbf{j}' = x(\mathbf{i}\cdot\mathbf{j}') + y(\mathbf{j}\cdot\mathbf{j}') + z(\mathbf{k}\cdot\mathbf{j}') \\ z' &= \mathbf{R}\cdot\mathbf{k}' = (x\mathbf{i} + y\mathbf{j} + z\mathbf{k})\cdot\mathbf{k}' = x(\mathbf{i}\cdot\mathbf{k}') + y(\mathbf{j}\cdot\mathbf{k}') + z(\mathbf{k}\cdot\mathbf{k}') \end{aligned}$$

From (1), the products $(\mathbf{i}\cdot\mathbf{i}')$, $(\mathbf{j}\cdot\mathbf{i}')$, . . . , $(\mathbf{k}\cdot\mathbf{k}')$ are just the cosines of the angles between the various axes of the two systems and are known from the data of the problem

When we consider products involving three rather than two vectors, we encounter the following possibilities·

$$(\mathbf{A}\cdot\mathbf{B})\mathbf{C} \qquad \mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) \qquad \mathbf{A}\times(\mathbf{B}\times\mathbf{C})$$

The first can be dismissed with a word In fact $\mathbf{A}\cdot\mathbf{B}$ is just a scalar, and thus $(\mathbf{A}\cdot\mathbf{B})\mathbf{C}$ is simply a vector whose length is $|\mathbf{A}\cdot\mathbf{B}|$ times the length of **C** and whose direction is the same as that of **C** or opposite to it, according as $\mathbf{A}\cdot\mathbf{B}$ is positive or negative.

For the product $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$, which is known as a **scalar triple product,** we observe first that the parentheses enclosing the vector product $\mathbf{B}\times\mathbf{C}$ are superfluous There is, in fact, only one alternative interpretation, namely, $(\mathbf{A}\cdot\mathbf{B})\times\mathbf{C}$, and this is meaningless, since both factors in a cross product must be vectors

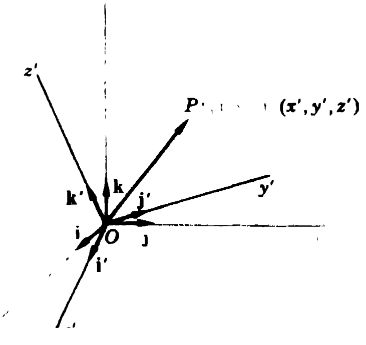


FIGURE 12.6 Two rectangular coordinate systems with the same origin.

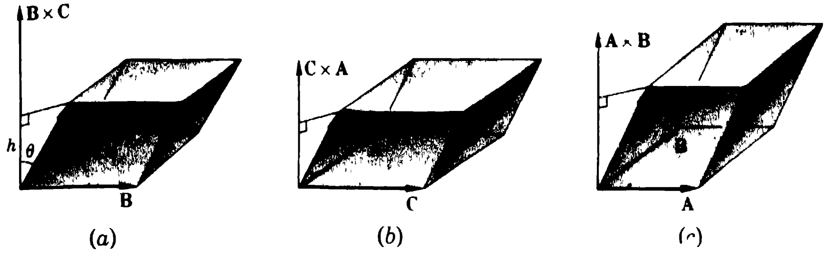


FIGURE 12.7
The geometrical interpretation of the scalar triple product

whereas **A** · **B** is a scalar Thus, no meaning but the intended one can be attached to the expression **A** · **B** × **C**, hence, it is customary to omit the parentheses

Geometrically, *the scalar triple product* **A** · **B** × **C** *represents the volume of the parallelepiped having the vectors* **A**, **B**, *and* **C** *as concurrent edges* For, if we regard the parallelogram having **B** and **C** as adjacent sides as the base of this figure, then **B** × **C** is a vector whose direction is perpendicular to the base and whose magnitude is equal to the area of the base Moreover, the altitude of the parallelepiped is the projection of **A** on **B** × **C** (Fig 12.7*a*) Hence, **A** · **B** × **C**, whose value is just the magnitude of **B** × **C** multiplied by the projection of **A** on **B** × **C**, is numerically equal to the volume of the parallelepiped If θ is less than $\pi/2$, i e., if **A** and **B** × **C** lie on the same side of the plane of **B** and **C**, then $\cos\theta$ is positive and so is the scalar triple product In particular, changing the order of the factors **B** and **C** gives the product **C** × **B**, whose direction, of course, is opposite to that of **B** × **C**; hence

(16) $$\mathbf{A}\cdot\mathbf{B}\times\mathbf{C} = -\mathbf{A}\cdot\mathbf{C}\times\mathbf{B}$$

Since the volume of a parallelepiped is independent of the face chosen as its base, it follows, by applying the preceding argument to Fig 12 7*b* and *c*, that **B** · **C** × **A** and **C** · **A** × **B** also give the volume of the same parallelepiped From this fact, together with (16), we therefore find

(17) $$\begin{aligned}\mathbf{A}\cdot\mathbf{B}\times\mathbf{C} &= \mathbf{B}\cdot\mathbf{C}\times\mathbf{A} = \mathbf{C}\cdot\mathbf{A}\times\mathbf{B}\\ &= -\mathbf{A}\cdot\mathbf{C}\times\mathbf{B} = -\mathbf{B}\cdot\mathbf{A}\times\mathbf{C} = -\mathbf{C}\cdot\mathbf{B}\times\mathbf{A}\end{aligned}$$

The first three arrangements can be obtained by starting anywhere on the circle in Fig 12 8 and reading the letters in the counterclockwise direction. For this reason they are said to be **cyclic permutations** of one another Similarly, the last three arrangements are cyclic permutations of one another which can be obtained by reading the letters in Fig 12 8 in the clockwise direction. Thus, (17) asserts that *any cyclic permutation of the factors in a scalar triple product leaves the value of the product unchanged, whereas a permutation which reverses the original cyclic order changes the sign of the product.*

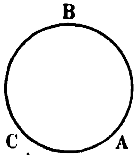


FIGURE 12.8 An illustration of cyclic and anticyclic permutations

Furthermore, since the order of factors in a dot product is immaterial, we find, by considering the first and third members of (17), that $\mathbf{A}\cdot\mathbf{B}\times\mathbf{C} = \mathbf{C}\cdot\mathbf{A}\times\mathbf{B} = \mathbf{A}\times\mathbf{B}\cdot\mathbf{C}$, which shows that *in any scalar triple product the dot and cross can be interchanged without altering the value of the product* For this reason it is customary to omit these symbols and write a scalar triple product simply as $[\mathbf{ABC}]$

If the vectors **A**, **B**, **C** all lie in the same plane or are parallel to the same plane, they necessarily form a parallelepiped of zero volume, and conversely Hence, $[\mathbf{ABC}] = 0$ is a necessary and sufficient condition that three vectors **A**, **B**, and **C** be parallel to one and the same plane In particular, if two factors of a scalar triple product have the same direction, the product is zero

Analytically, if we write

$$\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k} \qquad \mathbf{B} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k} \qquad \mathbf{C} = c_1\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k}$$

we have

$$\begin{aligned}\mathbf{A}\cdot\mathbf{B}\times\mathbf{C} &= (a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k})\cdot[(b_2c_3 - b_3c_2)\mathbf{i} - (b_1c_3 - b_3c_1)\mathbf{j} + (b_1c_2 - b_2c_1)\mathbf{k}]\\ &= a_1(b_2c_3 - b_3c_2) - a_2(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1)\end{aligned}$$

which is just the expanded form of the determinant

$$(18) \qquad [\mathbf{ABC}] = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

The relations in (17) are thus equivalent to the familiar fact that interchanging any two rows in a determinant changes the sign of the determinant

EXAMPLE 3

If **A**, **B**, and **C** are three vectors which are not parallel to the same plane, show that any vector **V** can be expressed as a linear combination of **A**, **B**, and **C**

If we write

$$(19) \qquad \mathbf{V} = a\mathbf{A} + b\mathbf{B} + c\mathbf{C}$$

where a, b, and c are scalar constants to be determined, and form the cross product of each member with the vector **B**, we obtain

$$\mathbf{V}\times\mathbf{B} = a\mathbf{A}\times\mathbf{B} + b\mathbf{B}\times\mathbf{B} + c\mathbf{C}\times\mathbf{B} = a\mathbf{A}\times\mathbf{B} + c\mathbf{C}\times\mathbf{B}$$

where the term $\mathbf{B}\times\mathbf{B}$ vanishes because its factors are identical. Now, if we form the dot product of the last result and the vector **C**, we have

$$\mathbf{V}\times\mathbf{B}\cdot\mathbf{C} = a\mathbf{A}\times\mathbf{B}\cdot\mathbf{C} + c\mathbf{C}\times\mathbf{B}\cdot\mathbf{C} = a\mathbf{A}\times\mathbf{B}\cdot\mathbf{C}$$

where the term $\mathbf{C} \times \mathbf{B} \cdot \mathbf{C}$ vanishes because it is a scalar triple product with two factors identical. By hypothesis, **A**, **B**, and **C** are not parallel to the same plane. Hence, $\mathbf{A} \times \mathbf{B} \cdot \mathbf{C}$ is different from zero, and we can solve for a, getting

$$a = \frac{[\mathbf{VBC}]}{[\mathbf{ABC}]}$$

In the same way we can obtain the remaining constants in the required linear combination

$$b = \frac{[\mathbf{AVC}]}{[\mathbf{ABC}]} \qquad c = \frac{[\mathbf{ABV}]}{[\mathbf{ABC}]}$$

Thus, under the conditions of the problem,

$$\mathbf{V} = \frac{[\mathbf{VBC}]}{[\mathbf{ABC}]}\mathbf{A} + \frac{[\mathbf{AVC}]}{[\mathbf{ABC}]}\mathbf{B} + \frac{[\mathbf{ABV}]}{[\mathbf{ABC}]}\mathbf{C} \tag{20}$$

The following special case of this result is often useful. If **V** is any vector parallel to the plane determined by **A** and **B**, then $[\mathbf{ABV}] = 0$ and the last term in the expansion (20) is zero. Hence it follows that, if **A** and **B** are vectors which are not parallel to the same line and if **V** is any vector parallel to the plane determined by **A** and **B**, then **V** can be expressed as a linear combination of **A** and **B**.

To express the **vector triple product** $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$ in a simpler expanded form, let us consider first the general case in which neither **A**, **B**, nor **C** is a zero vector and **B** and **C** are not parallel. Now, from the definition of a cross product it is clear that $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$ is a vector perpendicular to **A** and to $\mathbf{B} \times \mathbf{C}$. But $\mathbf{B} \times \mathbf{C}$ is itself perpendicular to the plane of **B** and **C**, and, thus, any vector such as $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$ which is perpendicular to $\mathbf{B} \times \mathbf{C}$ must lie in the plane of **B** and **C** (Fig. 12.9). Hence, by Example 3, the vector $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$ must be expressible as a linear combination of **B** and **C**, that is,

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \lambda\mathbf{B} + \mu\mathbf{C}$$

To find λ and μ, we first use the fact that $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$ is also perpendicular to **A** and, hence, that its dot product with **A** must be zero

$$\mathbf{A} \cdot [\mathbf{A} \times (\mathbf{B} \times \mathbf{C})] = \mathbf{A} \cdot (\lambda\mathbf{B} + \mu\mathbf{C}) = \lambda(\mathbf{A} \cdot \mathbf{B}) + \mu(\mathbf{A} \cdot \mathbf{C}) = 0$$

Thus,

$$\frac{\lambda}{\mu} = -\frac{\mathbf{A} \cdot \mathbf{C}}{\mathbf{A} \cdot \mathbf{B}}$$

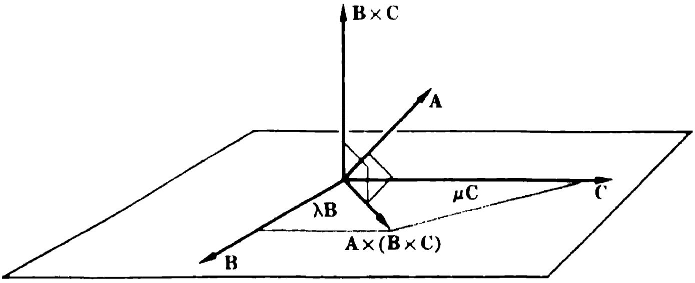


FIGURE 12.9
The geometrical interpretation of the vector triple product

and, therefore,

$$\lambda = \nu(\mathbf{A}\cdot\mathbf{C}) \qquad \mu = -\nu(\mathbf{A}\cdot\mathbf{B})$$

and

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = \nu[(\mathbf{A}\cdot\mathbf{C})\mathbf{B} - (\mathbf{A}\cdot\mathbf{B})\mathbf{C}] \tag{21}$$

To find ν, it is convenient to consider first the special case in which $\mathbf{A} = \mathbf{B}$.

$$\mathbf{B}\times(\mathbf{B}\times\mathbf{C}) = \nu_1[(\mathbf{B}\cdot\mathbf{C})\mathbf{B} - (\mathbf{B}\cdot\mathbf{B})\mathbf{C}] \tag{22}$$

Let θ be the angle between **B** and **C**, and form the dot product of **C** with each side of the last equality Then, using the properties of the scalar triple product and applying Formulas (1) and (2) to the resulting dot products, we have

$$\begin{aligned}\mathbf{B}\times(\mathbf{B}\times\mathbf{C})\cdot\mathbf{C} = -(\mathbf{B}\times\mathbf{C})\cdot(\mathbf{B}\times\mathbf{C}) &= \nu_1[(\mathbf{B}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{C}) - (\mathbf{B}\cdot\mathbf{B})(\mathbf{C}\cdot\mathbf{C})]\\ -|\mathbf{B}\times\mathbf{C}|^2 &= \nu_1(B^2C^2\cos^2\theta - B^2C^2)\\ -B^2C^2\sin^2\theta &= \nu_1B^2C^2(\cos^2\theta - 1) = -\nu_1B^2C^2\sin^2\theta\end{aligned}$$

Hence, $\nu_1 = 1$, and (22) becomes specificially

$$\mathbf{B}\times(\mathbf{B}\times\mathbf{C}) = (\mathbf{B}\cdot\mathbf{C})\mathbf{B} - (\mathbf{B}\cdot\mathbf{B})\mathbf{C} \tag{23}$$

We now return to Eq (21) and form the dot product of both members with **B**

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C})\cdot\mathbf{B} = \nu[(\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{B}) - (\mathbf{A}\cdot\mathbf{B})(\mathbf{C}.\mathbf{B})]$$

Now, by an obvious rearrangement of the scalar triple product on the left, we have

$$-\mathbf{A}\cdot\mathbf{B}\times(\mathbf{B}\times\mathbf{C}) = \nu[(\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{B}) - (\mathbf{A}\cdot\mathbf{B})(\mathbf{C}\cdot\mathbf{B})]$$

or, applying Eq (23) to the left-hand side,

$$-\mathbf{A}\cdot[(\mathbf{B}\cdot\mathbf{C})\mathbf{B} - (\mathbf{B}\cdot\mathbf{B})\mathbf{C}] = \nu[(\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{B}) - (\mathbf{A}\cdot\mathbf{B})(\mathbf{C}\cdot\mathbf{B})]$$

which will be true if and only if $\nu = 1$ † Hence, in general,

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = (\mathbf{A}\cdot\mathbf{C})\mathbf{B} - (\mathbf{A}\cdot\mathbf{B})\mathbf{C} \tag{24}$$

Moreover, if either **A**, **B**, or **C** is zero or if **B** and **C** have the same direction, it is evident by inspection that Eq (24) still holds Hence, the restrictions we imposed upon **A**, **B**, and **C** at the beginning of our discussion can be eliminated, and Eq. (24) is correct in all cases

By a straightforward application of Eq (24) we find that

$$(\mathbf{A}\times\mathbf{B})\times\mathbf{C} = -\mathbf{C}\times(\mathbf{A}\times\mathbf{B}) = -(\mathbf{C}\cdot\mathbf{B})\mathbf{A} + (\mathbf{C}\cdot\mathbf{A})\mathbf{B}$$

which is *not* equal to $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$ Hence, the position of the parentheses in a vector triple product is significant

With a knowledge of scalar and vector triple products, products involving more than three vectors can be expanded without difficulty For instance,

---

† Unless, of course, $(\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{B}) - (\mathbf{A}\cdot\mathbf{B})(\mathbf{C}\cdot\mathbf{B}) = 0$, in which case the value of ν is irrelevant.

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D})$$

can be regarded as the scalar triple product of the vectors **A**, **B**, and (**C** × **D**) This allows us to write

$$\begin{aligned}\mathbf{A} \times \mathbf{B} \cdot (\mathbf{C} \times \mathbf{D}) &= \mathbf{A} \cdot [\mathbf{B} \times (\mathbf{C} \times \mathbf{D})] \\ &= \mathbf{A} \cdot [(\mathbf{B} \cdot \mathbf{D})\mathbf{C} - (\mathbf{B} \cdot \mathbf{C})\mathbf{D}] \\ &= (\mathbf{A} \cdot \mathbf{C})(\mathbf{B} \cdot \mathbf{D}) - (\mathbf{A} \cdot \mathbf{D})(\mathbf{B} \cdot \mathbf{C})\end{aligned}$$

This result is sometimes referred to as **Lagrange's identity.**

Similarly, (**A** × **B**) × (**C** × **D**) can be thought of as the vector triple product of (**A** × **B**), **C**, and **D** or of **A**, **B**, and (**C** × **D**). Taking the former point of view and applying (24), we find

$$\begin{aligned}(\mathbf{A} \times \mathbf{B}) \times (\mathbf{C} \times \mathbf{D}) &= [\mathbf{A} \times \mathbf{B} \cdot \mathbf{D}]\mathbf{C} - [\mathbf{A} \times \mathbf{B} \cdot \mathbf{C}]\mathbf{D} \\ &= [\mathbf{ABD}]\mathbf{C} - [\mathbf{ABC}]\mathbf{D}\end{aligned}$$

which is a vector in the plane of **C** and **D** From the latter point of view,

$$\begin{aligned}(\mathbf{A} \times \mathbf{B}) \times (\mathbf{C} \times \mathbf{D}) &= -(\mathbf{C} \times \mathbf{D}) \times (\mathbf{A} \times \mathbf{B}) \\ &= -[\mathbf{C} \times \mathbf{D} \cdot \mathbf{B}]\mathbf{A} + [\mathbf{C} \times \mathbf{D} \cdot \mathbf{A}]\mathbf{B} \\ &= [\mathbf{CDA}]\mathbf{B} - [\mathbf{CDB}]\mathbf{A}\end{aligned}$$

which is a vector in the plane of **A** and **B** These two results together show that (**A** × **B**) × (**C** × **D**) is directed along the line of intersection of the plane of **A** and **B** and the plane of **C** and **D**

EXERCISES

1 For each of the following sets of vectors

a	$\mathbf{A} = 2\mathbf{i} - 2\mathbf{j} + \mathbf{k}$	b	$\mathbf{A} = 2\mathbf{i} - 3\mathbf{j} + 6\mathbf{k}$	c	$\mathbf{A} = 10\mathbf{i} + 10\mathbf{j} + 5\mathbf{k}$
	$\mathbf{B} = \mathbf{i} + 8\mathbf{j} - 4\mathbf{k}$		$\mathbf{B} = 10\mathbf{i} + 2\mathbf{j} + 11\mathbf{k}$		$\mathbf{B} = 5\mathbf{i} - 2\mathbf{j} - 14\mathbf{k}$
	$\mathbf{C} = 12\mathbf{i} - 4\mathbf{j} - 3\mathbf{k}$		$\mathbf{C} = 2\mathbf{i} - 9\mathbf{j} - 6\mathbf{k}$		$\mathbf{C} = 4\mathbf{i} + 7\mathbf{j} - 4\mathbf{k}$

what are the lengths of **A**, **B**, and **C**? What is **A** · **B**? **A** × **C**? the projection of **B** on **C**? the projection of **C** on **B**? the angle between **A** and **B**? [**ABC**]? **A** × (**B** × **C**)? the volume of the parallelepiped having **A** + **C**, **A** − **C**, and **B** as concurrent edges? the volume of the parallelepiped having **A** + **C**, **A** − **C**, and **C** as concurrent edges?

2 If **A**, **B**, and **C** are any three vectors, prove that

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) + \mathbf{B} \times (\mathbf{C} \times \mathbf{A}) + \mathbf{C} \times (\mathbf{A} \times \mathbf{B}) = 0$$

3 Prove that $(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) + (\mathbf{B} \times \mathbf{C}) \cdot (\mathbf{A} \times \mathbf{D}) + (\mathbf{C} \times \mathbf{A}) \cdot (\mathbf{B} \times \mathbf{D}) = 0$.

4 If the plane determined by **A** and **B** is perpendicular to the plane determined by **C** and **D**, show that $(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = 0$

5 Show that the volume of the tetrahedron having **A** + **B**, **B** + **C**, and **C** + **A** as concurrent edges is twice the volume of the tetrahedron having **A**, **B**, and **C** as concurrent edges

6 If **A** is a given vector and **X** · **A** = **Y** · **A**, can we conclude that **X** = **Y**?

7 Are two vectors equal if they have equal components in a given direction? in two given directions? in three given directions? in an arbitrary direction?

8 Find the unit vector perpendicular to both $\mathbf{i} - 2\mathbf{j} + \mathbf{k}$ and $3\mathbf{i} + \mathbf{j} - 2\mathbf{k}$.

9 Find the unit vector parallel to the plane of $\mathbf{i} + \mathbf{j} - 2\mathbf{k}$ and $3\mathbf{i} - 2\mathbf{j} + \mathbf{k}$ and perpendicular to $2\mathbf{i} + 2\mathbf{j} - \mathbf{k}$.

10 Show that, if the four vectors **A**, **B**, **C**, **D** are coplanar, then $(\mathbf{A}\times\mathbf{B})\times(\mathbf{C}\times\mathbf{D}) = 0$. Is the converse true?

11 Show that, if $\mathbf{A}+\mathbf{B}+\mathbf{C} = 0$, then $\mathbf{A}\times\mathbf{B} = \mathbf{B}\times\mathbf{C} = \mathbf{C}\times\mathbf{A}$. Is the converse true?

12 Prove that two nonzero vectors are linearly dependent if and only if they are parallel.

13 Prove that three vectors are linearly dependent if and only if they are parallel to the same plane.

14 Prove that four vectors are always linearly dependent. [Hint: Expand $(\mathbf{A}\times\mathbf{B})\times(\mathbf{C}\times\mathbf{D})$ in two different ways and equate the results.]

15 Prove that, for all values of the a's and b's,

$$(a_1b_1 + a_2b_2 + a_3b_3)^2 \leqq (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2)$$

This is the special case $n = 3$ of what is known as **Cauchy's inequality**,

$$\left(\sum_{i=1}^{n} a_ib_i\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right)$$

16 By considering the dot product of the two vectors $\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j}$ and $\mathbf{B} = b_1\mathbf{i} + b_2\mathbf{j}$, derive the formula for the cosine of the difference of two angles.

17 By considering the cross product of the two vectors of Exercise 16, derive the formula for the sine of the difference of two angles.

18 Show that, if $\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ is a constant vector drawn from the origin, the locus of the end points of the vectors $\mathbf{R} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ which satisfy the equation $(\mathbf{R}-\mathbf{A})\cdot\mathbf{A} = 0$ is a plane perpendicular to **A** at its end point. What is the locus of the end points of the vectors which satisfy the equation $(\mathbf{R}-\mathbf{A})\cdot\mathbf{R} = 0$? $(\mathbf{R}-\mathbf{A})\cdot(\mathbf{R}-\mathbf{A}) = 0$?

19 The three noncollinear points L, M, and N lie in the plane p. Prove that, if **L**, **M**, and **N** are the vectors to these points from an origin not lying in p, the vector

$$(\mathbf{L}\times\mathbf{M}) + (\mathbf{M}\times\mathbf{N}) + (\mathbf{N}\times\mathbf{L})$$

is perpendicular to p.

20 Carry through in detail the geometrical proof that dot multiplication is distributive over addition.

21 Carry through in detail the geometrical proof that cross multiplication is distributive over addition.

22 Prove that $(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{B}\times\mathbf{C})\times(\mathbf{C}\times\mathbf{A}) = [\mathbf{ABC}]^2$.

23 If **A**, **B**, and **C** are any three independent vectors, the vectors

$$\mathbf{U} = \frac{\mathbf{B}\times\mathbf{C}}{[\mathbf{ABC}]} \qquad \mathbf{V} = \frac{\mathbf{C}\times\mathbf{A}}{[\mathbf{ABC}]} \qquad \mathbf{W} = \frac{\mathbf{A}\times\mathbf{B}}{[\mathbf{ABC}]}$$

are said to form a set **reciprocal** to the set **A**, **B**, **C**. Show that

$$\mathbf{A}\cdot\mathbf{U} = \mathbf{B}\cdot\mathbf{V} = \mathbf{C}\cdot\mathbf{W} = 1 \qquad \text{and that} \qquad [\mathbf{UVW}] = 1/[\mathbf{ABC}]$$

If $\mathbf{A} = \mathbf{i} + 2\mathbf{j} - 2\mathbf{k}$, $\mathbf{B} = \mathbf{i} + 8\mathbf{j} + 4\mathbf{k}$, and $\mathbf{C} = 12\mathbf{i} - 4\mathbf{j} + 3\mathbf{k}$, express the vector $\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$ as a linear combination of **A**, **B**, and **C** and as a linear combination of the vectors **U**, **V**, and **W** of the set reciprocal to **A**, **B**, and **C**.

24 Show that, if $\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$, $\mathbf{B} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$, $\mathbf{C} = c_1\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k}$, and $D = d_1\mathbf{i} + d_2\mathbf{j} + d_3\mathbf{k}$, then the system of equations

$$\begin{aligned} a_1x + b_1y + c_1z &= d_1 \\ a_2x + b_2y + c_2z &= d_2 \\ a_3x + b_3y + c_3z &= d_3 \end{aligned}$$

is equivalent to the single vector equation $x\mathbf{A} + y\mathbf{B} + z\mathbf{C} = \mathbf{D}$. Assuming that $[\mathbf{ABC}] \neq 0$, solve this vector equation for x, y, and z, and show that the result is equivalent to that obtained from the algebraic form of the system by using Cramer's rule (Theorem 7, Sec. 10.3).

25 In mechanics the **moment** M **of a force F about a point** O is defined as the magnitude of **F** times the perpendicular distance from the point O to the line of action of **F**. If the vector **moment M** is defined as the vector whose magnitude is M and whose direction is perpendicular to the plane of O and **F**, show that $\mathbf{M} = \mathbf{R} \times \mathbf{F}$, where **R** is the vector from O to any point on the line of action of **F**. Would $\mathbf{M} = \mathbf{F} \times \mathbf{R}$ be equally acceptable? Explain

12.2

Vector functions of one variable

If t is a scalar variable and if to each value of t in some range there corresponds a value of a vector **V**, we say that **V** is a **vector function of** t. Since the component of a vector in any direction is known whenever the vector itself is known, it follows that, if **V** is a function of t, so, too, are its components in the directions of the unit vectors **i**, **j** and **k**. Hence, we can write

(1) $$\mathbf{V}(t) = V_1(t)\mathbf{i} + V_2(t)\mathbf{j} + V_3(t)\mathbf{k}$$

In particular, we say that $\mathbf{V}(t)$ is **continuous** if and only if the three scalar functions $V_1(t)$, $V_2(t)$, and $V_3(t)$ are continuous

If the independent variable t of a vector function $\mathbf{V}(t)$ changes by an amount Δt, the function will in general change both in magnitude and in direction. In other words, corresponding to the scalar increment Δt we have the vector increment

$$\begin{aligned}\Delta\mathbf{V} &= \mathbf{V}(t + \Delta t) - \mathbf{V}(t)\\ &= [V_1(t + \Delta t)\mathbf{i} + V_2(t + \Delta t)\mathbf{j} + V_3(t + \Delta t)\mathbf{k}]\\ &\qquad\qquad - [V_1(t)\mathbf{i} + V_2(t)\mathbf{j} + V_3(t)\mathbf{k}]\end{aligned}$$

(2) $$= \Delta V_1\mathbf{i} + \Delta V_2\mathbf{j} + \Delta V_3\mathbf{k}$$

By the **derivative of a vector function** $\mathbf{V}(t)$, we mean, as usual,

$$\frac{d\mathbf{V}}{dt} = \lim_{\Delta t\to 0} \frac{\mathbf{V}(t + \Delta t) - \mathbf{V}(t)}{\Delta t} = \lim_{\Delta t\to 0} \frac{\Delta\mathbf{V}}{\Delta t}$$

or, using (2),

$$\frac{d\mathbf{V}}{dt} = \lim_{\Delta t\to 0} \frac{\Delta V_1}{\Delta t}\mathbf{i} + \lim_{\Delta t\to 0} \frac{\Delta V_2}{\Delta t}\mathbf{j} + \lim_{\Delta t\to 0} \frac{\Delta V_3}{\Delta t}\mathbf{k}$$

(3) $$= \frac{dV_1}{dt}\mathbf{i} + \frac{dV_2}{dt}\mathbf{j} + \frac{dV_3}{dt}\mathbf{k}$$

From (3) we are motivated to define the **differential of a vector function** $\mathbf{V}(t)$ to be

(4) $$d\mathbf{V} = dV_1\mathbf{i} + dV_2\mathbf{j} + dV_3\mathbf{k}$$

In particular, for the very important vector

(5) $$\mathbf{R} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

drawn from the origin to the point (x,y,z), we have

(6) $$d\mathbf{R} = dx\,\mathbf{i} + dy\,\mathbf{j} + dz\,\mathbf{k}$$

From the definition of the derivative of a vector function of one variable it follows that sums, differences, and products of vectors can be differentiated by formulas just like those of ordinary calculus, provided that the proper order of factors is maintained wherever the order is significant. Specifically, we have

$$\frac{d(\mathbf{U} \pm \mathbf{V})}{dt} = \frac{d\mathbf{U}}{dt} \pm \frac{d\mathbf{V}}{dt} \tag{7}$$

$$\frac{d(\phi\mathbf{V})}{dt} = \frac{d\phi}{dt}\mathbf{V} + \phi\frac{d\mathbf{V}}{dt} \tag{8}$$

$$\frac{d(\mathbf{U} \cdot \mathbf{V})}{dt} = \frac{d\mathbf{U}}{dt} \cdot \mathbf{V} + \mathbf{U} \cdot \frac{d\mathbf{V}}{dt} \tag{9}$$

$$\frac{d(\mathbf{U} \times \mathbf{V})}{dt} = \frac{d\mathbf{U}}{dt} \times \mathbf{V} + \mathbf{U} \times \frac{d\mathbf{V}}{dt} \tag{10}$$

$$\frac{d[\mathbf{UVW}]}{dt} = \left[\frac{d\mathbf{U}}{dt}\mathbf{VW}\right] + \left[\mathbf{U}\frac{d\mathbf{V}}{dt}\mathbf{W}\right] + \left[\mathbf{UV}\frac{d\mathbf{W}}{dt}\right] \tag{11}$$

$$\frac{d[\mathbf{U} \times (\mathbf{V} \times \mathbf{W})]}{dt} = \frac{d\mathbf{U}}{dt} \times (\mathbf{V} \times \mathbf{W}) + \mathbf{U} \times \left(\frac{d\mathbf{V}}{dt} \times \mathbf{W}\right) + \mathbf{U} \times \left(\mathbf{V} \times \frac{d\mathbf{W}}{dt}\right) \tag{12}$$

The simplest example of a vector function of one variable is the set of vectors drawn from the origin to the points of a curve C on which the scalar variable t is a parameter For a general point on C is associated with a unique value of the parameter, say $t = t_1$, and determines with the origin a unique vector $\mathbf{V}(t_1)$ (Fig 12 10a). This correspondence between the values of t and the vectors $\mathbf{V}(t)$ is clearly a vector function of t according to our definition. Conversely, if the values of a continuous vector function $\mathbf{V}(t)$ are drawn from a common origin, their end points will define a curve C whose points will be in correspondence with the values of the scalar variable t

This point of view leads to an important geometric interpretation of the derivative $d\mathbf{V}/dt$ For, since Δt is just a scalar, the quotient $\Delta\mathbf{V}/\Delta t$ is a well-defined vector having the same direction as $\Delta\mathbf{V}$ itself Moreover, as Fig. 12 10b shows, the direction of $\Delta\mathbf{V}$ is that of an infinitesimal chord of the curve C Therefore, as Δt approaches 0, the direction of $\Delta\mathbf{V}$, and, hence, the direction of $\Delta\mathbf{V}/\Delta t$, approaches the direction of a tangent to C That is, $d\mathbf{V}/dt$

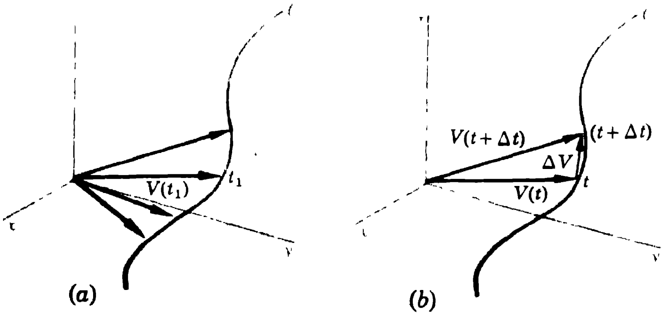


FIGURE 12 10 The geometrical interpretation of a vector function of one variable

is a vector tangent to the curve C which is the locus of the end points of the vectors **V**(t) In particular, if the scalar variable t is taken to be the arc length s of C, measured from some reference point on C, we have

$$\left|\frac{d\mathbf{V}}{ds}\right| = \lim_{\Delta s\to 0}\left|\frac{\Delta\mathbf{V}}{\Delta s}\right| = \lim_{\Delta s\to 0}\frac{\text{infinitesimal chord of } C}{\text{infinitesimal arc of } C} = 1$$

Hence, ***if s is the arc length of the curve C defined by the end points of the vectors* V*(s), then d*V*/ds is a unit vector tangent to C.***

EXAMPLE 1

At what point or points is the tangent to the curve $x = t^3$, $y = 5t^2$, $z = 10t$ perpendicular to the tangent at the point where $t = 1$?

From our earlier discussion it is clear that the given curve is equivalent to the vector function

$$\mathbf{V}(t) = t^3\mathbf{i} + 5t^2\mathbf{j} + 10t\mathbf{k}$$

Moreover, the tangent to this curve at a general point t is

$$\frac{d\mathbf{V}}{dt} = 3t^2\mathbf{i} + 10t\mathbf{j} + 10\mathbf{k}$$

and, in particular, at $t = 1$ the tangent is

$$3\mathbf{i} + 10\mathbf{j} + 10\mathbf{k}$$

Using the fact that two nonzero vectors are perpendicular if and only if their dot product vanishes, it follows that the tangent at a general point t will be perpendicular to the tangent at the point $t = 1$ if and only if

$$3(3t^2) + 10(10t) + 10(10) = 9t^2 + 100t + 100 = 0$$

This condition holds for the two values $t = -10/9$, -10 Hence, evaluating the x, y, and z coordinates of the points with these parameters, it follows that the tangent at

$$\left(-\frac{1{,}000}{729}, \frac{500}{81}, -\frac{100}{9}\right)$$

and the tangent at $(-1{,}000, 500, -100)$ are both perpendicular to the tangent at $t = 1$ and that these are the only points with this property

EXAMPLE 2

Discuss from the point of view of vector analysis the problem of the determination of the velocity and acceleration of a particle moving along a curve C

To do this, let us suppose that the path C, which is the locus of the instantaneous positions of the moving particle, is defined by the vector function **P**(t), where t is the time In other words, **P**(t) is the vector drawn from the origin to the position of the moving particle at the general time t

Now let s be the arc length of C Then, by the chain rule, we can write

$$\frac{d\mathbf{P}}{dt} = \frac{d\mathbf{P}}{ds}\frac{ds}{dt} \tag{13}$$

Since ds/dt is the speed v of the moving particle and since $d\mathbf{P}/ds$ is a unit vector tangent to the path of the particle, it follows from (13) that the vector

$$\mathbf{v} = \frac{d\mathbf{P}}{dt} \tag{14}$$

agrees both in magnitude and in direction with the velocity of the particle and, thus, can properly be called its vector velocity.

Moreover, if we define the **vector acceleration** of the particle as the time derivative of its vector velocity and, for convenience, denote the general unit vector tangent to C, namely, $d\mathbf{P}/ds$, by the symbol $\mathbf{T}$, so that (14) becomes

$$\mathbf{v} = v\mathbf{T}$$

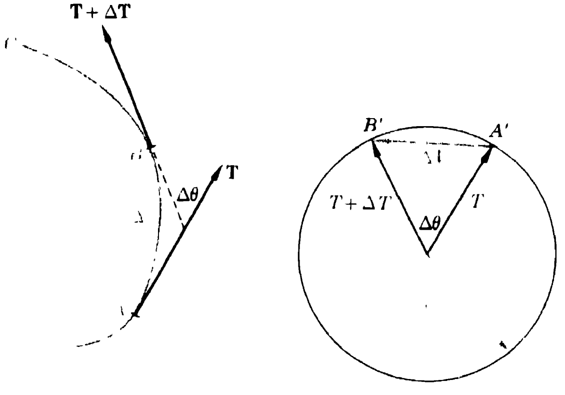
we can write

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d(v\mathbf{T})}{dt} = \frac{dv}{dt}\mathbf{T} + v\frac{d\mathbf{T}}{dt} = \frac{dv}{dt}\mathbf{T} + v\frac{d\mathbf{T}}{ds}\frac{ds}{dt}$$

$$(15) \qquad = \frac{dv}{dt}\mathbf{T} + v^2\frac{d\mathbf{T}}{ds}$$

In the first term on the right in (15) the scalar quantity dv/dt is the rate of change of the tangential speed v. Therefore, since $\mathbf{T}$ is by definition a unit vector tangent to C, the product $(dv/dt)\mathbf{T}$ is in magnitude and direction just the **tangential acceleration** of the moving particle.

To interpret the second term on the right in (15), we observe that, since $\mathbf{T}$ is a unit vector, it can vary only in direction. Hence, if the various values of $\mathbf{T}$ are drawn from a common origin, the locus of their end points will be a curve on a sphere of unit radius. Now the length of the increment $\Delta\mathbf{T}$ (Fig. 12.11*b*) is approximately the length of the arc $A'B'$, which in turn is equal to $\Delta\theta$, where $\Delta\theta$ is the angle between the tangents to C at the points A and B, a distance Δs apart (Fig. 12.11*a*). Hence,



(*a*) (*b*)

FIGURE 12.11 The unit tangents to a space curve plotted from a common origin.

$$\left|\frac{d\mathbf{T}}{ds}\right| = \lim_{\Delta s\to 0}\frac{|\Delta\mathbf{T}|}{|\Delta s|} = \lim_{\Delta s\to 0}\frac{\text{angle between tangents to } C \text{ at } A \text{ and } B}{\text{arc length along } C \text{ between } A \text{ and } B}$$

$$= \text{curvature of } C$$

Moreover, from Fig. 12.11*b* it is evident that the limiting direction of $\Delta\mathbf{T}$ is perpendicular to $\mathbf{T}$ in the plane which $\mathbf{T}$ and $\mathbf{T} + \Delta\mathbf{T}$ determine in the limit. If C is a plane curve, this, of course, is the unique plane in which C lies. If C is a twisted curve, this plane, which is known as the **osculating plane,** will vary from point to point along C. Hence, to summarize, $d\mathbf{T}/ds$ is a vector perpendicular to C in the osculating plane of C whose magnitude is equal to the curvature K of C.

If, finally, we let $\mathbf{N}$ denote a unit normal drawn toward the concave side of C in the osculating plane and define the radius of curvature of C as

$$\rho = \frac{1}{K}$$

we can write Eq (15) in the form

$$\mathbf{a} = \frac{dv}{dt}\mathbf{T} + \frac{v^2}{\rho}\mathbf{N} \tag{16}$$

which shows that *at any point in its path, the vector acceleration of a moving particle is the sum of a component of magnitude dv/dt along the tangent to the path and a component of magnitude v^2/ρ normal to the path in the osculating plane to the path*

EXERCISES

1 If $\mathbf{P} = \mathbf{A}\cos kt + \mathbf{B}\sin kt$, where $\mathbf{A}$ and $\mathbf{B}$ are arbitrary constant vectors, show that $\mathbf{P} \times d\mathbf{P}/dt$ is a constant and that $d^2\mathbf{P}/dt^2 + k^2\mathbf{P} = 0$

2 If $\mathbf{P}$ is any vector, show that $\dfrac{d}{dt}\left(\mathbf{P} \times \dfrac{d\mathbf{P}}{dt}\right) = \mathbf{P} \times \dfrac{d^2\mathbf{P}}{dt^2}$.

3 What is the derivative (a) of $\mathbf{U} \cdot \dfrac{d\mathbf{U}}{dt} \times \dfrac{d^2\mathbf{U}}{dt^2}$? (b) of $\mathbf{U} \times \left(\dfrac{d\mathbf{U}}{dt} \times \dfrac{d^2\mathbf{U}}{dt^2}\right)$?

4 If $\mathbf{V}$ is an arbitrary vector function of t, is $|d\mathbf{V}| = d|\mathbf{V}|$?

5 If $\mathbf{V}$ is an arbitrary function of t, show that $\mathbf{V} \cdot \dfrac{d\mathbf{V}}{dt} = V\dfrac{dV}{dt}$

6 What is the angle between the tangents to the curve $x = t$, $y = t^2$, $z = t^3$ at the points where $t = 1$ and $t = -1$?

7 Show that there are no pairs of points on the curve $x = t$, $y = t^2$, $z = t^3$ at which the tangents are parallel Are there such pairs of points on the curve $x = 3t^4 - 6t^2 + 12t$, $y = 4t^3 - 6t^2$, $z = 12t^2$ on the curve $x = 15t$, $y = 5t^3$, $z = 15t + 3t^5$?

8 If $\mathbf{R} = t^2\mathbf{i} \quad t^3\mathbf{j} + t^4\mathbf{k}$ is the vector from the origin to a moving particle, find the resultant velocity of the particle when $t = 1$ What is the component of this velocity in the direction of the vector $8\mathbf{i} - \mathbf{j} + 4\mathbf{k}$? What is the vector acceleration of the particle? What are the tangential and normal components of its acceleration?

9 If a particle starts to move from rest at the point $(0,1,2)$ with component accelerations $a_x = 1 + 2t$, $a_y = t^3$, $a_z = 2t - t^2$, find the vector from the origin to the instantaneous position of the particle

10 If $\mathbf{R}_1, \mathbf{R}_2, \quad , \mathbf{R}_n$ are the vectors from the origin to the respective mass particles $m_1, m_2, \quad , m_n$, the end point of the vector

$$\mathbf{C} = \frac{\sum_{i=1}^{n} m_i\mathbf{R}_i}{\sum_{i=1}^{n} m_i}$$

is called the **center of gravity** of the system of particles Show that, for any vector $\mathbf{R}$,

$$\sum_{i=1}^{n} m_i(\mathbf{R} - \mathbf{R}_i) \cdot (\mathbf{R} - \mathbf{R}_i) = m(\mathbf{R} - \mathbf{C}) \quad (\mathbf{R} - \mathbf{C}) + \sum_{i=1}^{n} m(\mathbf{C} - \mathbf{R}_i) \cdot (\mathbf{C} - \mathbf{R}_i)$$

where m is the total mass of all the particles

11 If a particle moves under the influence of a force $\mathbf{F}$ which is always directed toward the origin, show that $\mathbf{R} \times d^2\mathbf{R}/dt^2 = 0$, where $\mathbf{R}$ is the vector from the origin to the particle (Hint· Newton's law, i e , mass × acceleration = force, remains correct when the acceleration and the force are interpreted as vector quantities)

12 If $\mathbf{R}(t)$ is the vector from the origin to the instantaneous position of a particle moving along a curve C, show that $\mathbf{R} \times d\mathbf{R}$ is equal to twice the area of the sector defined by the two vectors $\mathbf{R}(t)$ and $\mathbf{R}(t + dt) \cong \mathbf{R} + d\mathbf{R}$ and the arc of C which they intercept Hence show that (a) if $\mathbf{R} \times d\mathbf{R}/dt = 0$, the vector $\mathbf{R}$ has a constant direction, and that (b) if $\mathbf{R} \times$

$d^2\mathbf{R}/dt^2 = 0$, the particle moves so that the radius vector **R** sweeps out equal areas in equal times. [Property **b** is a generalization of one of the laws of planetary motion discovered by Johannes Kepler (1571–1630).]

13 If **T** is a unit vector tangent to C and if **N** is the unit normal to C in the osculating plane, the vector $\mathbf{B} = \mathbf{T} \times \mathbf{N}$ is called the **binormal** to C at the point where **T** and **N** are drawn. Using the fact that $d\mathbf{T}/ds = \mathbf{N}/\rho$, show that $d\mathbf{B}/ds = \mathbf{T} \times d\mathbf{N}/ds$ and hence that $d\mathbf{B}/ds$ has the same direction as **N**. (The absolute value of $d\mathbf{B}/ds$ is called the **torsion** of the curve C and measures the rate at which the osculating plane turns as we move along C.)

14 What is the equation of the osculating plane to the space curve $x = t^4$, $y = t^2$, $z = t^3$ at the point P_1 (x_1,y_1,z_1) whose parameter is $t = t_1$? [Hint: Let P (x,y,z) be a general point in the osculating plane, and impose the condition that the vector joining P to P_1 be coplanar with the vectors **T** and $d\mathbf{T}/dt$ at P_1.]

15 What is the osculating plane to the curve $x = t$, $y = t^2$, $z = t^3$ at the point whose parameter is t? What is the normal to this curve at the point $t = 1$? What is the binormal at $t = 1$?

12.3

The operator ∇

Let $\phi(x,y,z)$ be a scalar function of position possessing first partial derivatives with respect to x, y, and z throughout some region of space, and let $\mathbf{R} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ be the vector drawn from the origin to a general point P (x,y,z). If we move from P to a neighboring point Q $(x + \Delta x, y + \Delta y, z + \Delta z)$ (Fig. 12.12), the function ϕ will change by an amount $\Delta\phi$ whose exact value, as derived in calculus, is

$$\Delta\phi = \frac{\partial\phi}{\partial x}\Delta x + \frac{\partial\phi}{\partial y}\Delta y + \frac{\partial\phi}{\partial z}\Delta z + \epsilon_1\,\Delta x + \epsilon_2\,\Delta y + \epsilon_3\,\Delta z \tag{1}$$

where ϵ_1, ϵ_2, ϵ_3 are quantities which approach zero as Q approaches P, i.e., as Δx, Δy, and Δz approach zero. If we divide the change $\Delta\phi$ by the distance $\Delta s \equiv |\Delta\mathbf{R}|$ between P and Q, we obtain a measure of the rate at which ϕ changes when we move from P to Q:

$$\frac{\Delta\phi}{\Delta s} = \frac{\partial\phi}{\partial x}\frac{\Delta x}{\Delta s} + \frac{\partial\phi}{\partial y}\frac{\Delta y}{\Delta s} + \frac{\partial\phi}{\partial z}\frac{\Delta z}{\Delta s} + \epsilon_1\frac{\Delta x}{\Delta s} + \epsilon_2\frac{\Delta y}{\Delta s} + \epsilon_3\frac{\Delta z}{\Delta s} \tag{2}$$

For instance, if $\phi(x,y,z)$ is the temperature at the general point P (x,y,z) then $\Delta\phi/\Delta s$ is the average rate of change of temperature in degrees per unit length at the point P in the direction in which Δs is measured. The limiting value of $\Delta\phi/\Delta s$ as Q approaches P

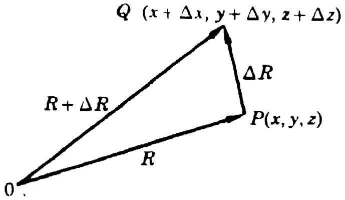


FIGURE 12.12 The coordinate vectors of two neighboring points

along the segment PQ is called the **derivative of ϕ in the direction PQ** or simply the **directional derivative** of ϕ. Clearly, in the limit the last three terms in (2) become zero and we have explicitly

$$\frac{d\phi}{ds} = \frac{\partial\phi}{\partial x}\frac{dx}{ds} + \frac{\partial\phi}{\partial y}\frac{dy}{ds} + \frac{\partial\phi}{\partial z}\frac{dz}{ds} \tag{3}$$

The first factor in each product on the right in (3) depends only on ϕ and the coordinates of the point P at which the derivatives of ϕ are evaluated. The second factor in each product is independent of ϕ and depends only on the direction in which the derivative is being computed. This observation suggests that $d\phi/ds$ can be thought of as the dot product of two vectors, one depending only on ϕ and the coordinates of P, the other depending only on the direction of ds, and in fact we can write

$$\begin{aligned}\frac{d\phi}{ds} &= \left(\frac{\partial\phi}{\partial x}\mathbf{i} + \frac{\partial\phi}{\partial y}\mathbf{j} + \frac{\partial\phi}{\partial z}\mathbf{k}\right)\cdot\left(\frac{dx}{ds}\mathbf{i} + \frac{dy}{ds}\mathbf{j} + \frac{dz}{ds}\mathbf{k}\right)\\ &= \left(\frac{\partial\phi}{\partial x}\mathbf{i} + \frac{\partial\phi}{\partial y}\mathbf{j} + \frac{\partial\phi}{\partial z}\mathbf{k}\right)\cdot\frac{d\mathbf{R}}{ds}\end{aligned} \tag{4}$$

The vector function

$$\frac{\partial\phi}{\partial x}\mathbf{i} + \frac{\partial\phi}{\partial y}\mathbf{j} + \frac{\partial\phi}{\partial z}\mathbf{k}$$

is known as the **gradient** of ϕ or simply **grad ϕ**, and in this notation (4) can be rewritten in the form

$$\frac{d\phi}{ds} = (\mathbf{grad}\ \phi)\cdot\frac{d\mathbf{R}}{ds} \tag{4a}$$

To determine the significance of **grad ϕ**, we observe first that, since Δs is by definition just the length of $\Delta\mathbf{R}$, it follows that $d\mathbf{R}/ds$ is a unit vector. Hence the dot product $(\mathbf{grad}\ \phi)\cdot(d\mathbf{R}/ds)$ is just the projection of **grad ϕ** in the direction of $d\mathbf{R}/ds$. Thus, according to (4a), **grad ϕ** *has the property that its projection in any direction is equal to the derivative of ϕ in that direction* (Fig. 12.13a). Since the maximum projection of a vector is the vector itself, it is clear that **grad ϕ** *extends in the direction of the greatest rate of change of ϕ and has that rate of change for its length.*

If we set $\varphi(x,y,z) = c$, we obtain, as c takes on different

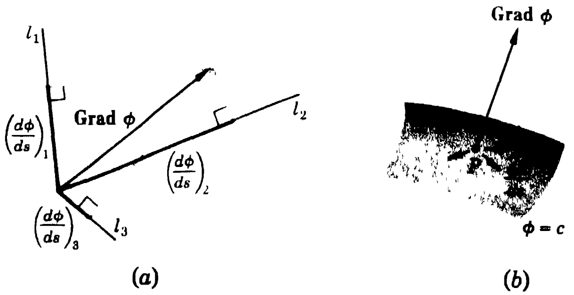

l_1
Grad ϕ
l_2
$\left(\frac{d\phi}{ds}\right)_1$
$\left(\frac{d\phi}{ds}\right)_2$
$\left(\frac{d\phi}{ds}\right)_3$
l_3
(a)
Grad ϕ
$\phi = c$
(b)

FIGURE 12.13 The geometrical interpretation of the gradient.

values, a family of surfaces known as the **level surfaces*** of ϕ, and, on the assumption that ϕ is a single-valued function, one and only one level surface passes through any given point P. If we now consider the level surface of ϕ which passes through P and fix our attention on neighboring points Q which lie on this surface, we have $\frac{\Delta\phi}{\Delta s} = 0$, since $\Delta\phi = 0$ because, by definition, ϕ has the same value at all points of a level surface. Hence, by (4*a*),

(5) $$(\mathbf{grad}\ \phi)\cdot\frac{d\mathbf{R}}{ds} = 0$$

for any vector $d\mathbf{R}/ds$ which has the limiting direction of a secant PQ of the level surface. Clearly, such vectors are all tangent to $\phi = c$ at the point P; hence, from the vanishing of the dot product in (5) it follows that **grad** ϕ is perpendicular to every tangent to the level surface at P. In other words, *the gradient of ϕ at any point P is perpendicular to the level surface of ϕ which passes through that point* (Fig. 12.13*b*). Evidently, **grad** ϕ is related to the level surfaces of ϕ in a way which is independent of the particular coordinate system used to describe ϕ. In other words, **grad** ϕ depends only on the intrinsic properties of ϕ. It follows, therefore that in the expression

$$\mathbf{grad}\ \phi = \frac{\partial\phi}{\partial x}\mathbf{i} + \frac{\partial\phi}{\partial y}\mathbf{j} + \frac{\partial\phi}{\partial z}\mathbf{k}$$

i, **j**, and **k** can be replaced by any other set of mutually perpendicular unit vectors provided that $\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z}$ are replaced by the directional derivatives of ϕ along the new axes.

The gradient of a function is frequently written in operational form as

$$\mathbf{grad}\ \phi = \left(\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\right)\phi$$

The operational "vector" thus defined is usually denoted by the symbol ∇ (read "del")

(6) $$\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}$$

In this notation our earlier results can be written

(7) $$\mathbf{grad}\ \phi = \nabla\phi$$

(8) $$\frac{d\phi}{ds} = \nabla\phi\cdot\frac{d\mathbf{R}}{ds}$$

(9) $$d\phi = \nabla\phi\cdot d\mathbf{R}$$

---

* This name, which is used regardless of the number of independent variables, is suggested by the analogy between the general case and the two-dimensional topographic interpretation in which $\phi(x,y)$ is the elevation at the point (x,y) and the loci $\phi(x,y) = c$ are the contour lines, i.e., curves consisting of points where the elevation above (or below) the xy-plane is constant.

Also, if ϕ is a function of u and u is a function of x, y, and z, then

$$\begin{aligned}
\nabla\phi &= \frac{\partial\phi}{\partial x}\mathbf{i} + \frac{\partial\phi}{\partial y}\mathbf{j} + \frac{\partial\phi}{\partial z}\mathbf{k} \\
&= \frac{d\phi}{du}\frac{\partial u}{\partial x}\mathbf{i} + \frac{d\phi}{du}\frac{\partial u}{\partial y}\mathbf{j} + \frac{d\phi}{du}\frac{\partial u}{\partial z}\mathbf{k} \\
&= \frac{d\phi}{du}\left(\frac{\partial u}{\partial x}\mathbf{i} + \frac{\partial u}{\partial y}\mathbf{j} + \frac{\partial u}{\partial z}\mathbf{k}\right) \\
&= \frac{d\phi}{du}\nabla u \qquad (10)
\end{aligned}$$

EXAMPLE 1

What is the directional derivative of the function $\phi(x,y,z) = xy^2 + yz^3$ at the point $(2,-1,1)$ in the direction of the vector $\mathbf{i} + 2\mathbf{j} + 2\mathbf{k}$?

Our first step must be to find the gradient of ϕ at the point $(2,-1,1)$. This is

$$\nabla\phi = \frac{\partial(xy^2 + yz^3)}{\partial x}\mathbf{i} + \frac{\partial(xy^2 + yz^3)}{\partial y}\mathbf{j} + \frac{\partial(xy^2 + yz^3)}{\partial z}\mathbf{k}\,\bigg|_{2,-1,1}$$

$$= y^2\mathbf{i} + (2xy + z^3)\mathbf{j} + 3yz^2\mathbf{k}\,\bigg|_{2,-1,1}$$

$$= \mathbf{i} - 3\mathbf{j} - 3\mathbf{k}$$

The projection of this in the direction of the given vector will be the required directional derivative. Since this projection can be found at once as the dot product of $\nabla\phi$ and a unit vector in the given direction, we next reduce $\mathbf{i} + 2\mathbf{j} + 2\mathbf{k}$ to a unit vector by dividing it by its magnitude, getting

$$\frac{\mathbf{i} + 2\mathbf{j} + 2\mathbf{k}}{\sqrt{1 + 4 + 4}} = \frac{1}{3}\mathbf{i} + \frac{2}{3}\mathbf{j} + \frac{2}{3}\mathbf{k}$$

The answer to our problem is, therefore,

$$\nabla\phi \cdot (\tfrac{1}{3}\mathbf{i} + \tfrac{2}{3}\mathbf{j} + \tfrac{2}{3}\mathbf{k}) = (\mathbf{i} - 3\mathbf{j} - 3\mathbf{k}) \cdot (\tfrac{1}{3}\mathbf{i} + \tfrac{2}{3}\mathbf{j} + \tfrac{2}{3}\mathbf{k}) = -\tfrac{11}{3}$$

The negative sign, of course, indicates that ϕ decreases in the given direction.

EXAMPLE 2

What is the unit normal to the surface $xy^3z^2 = 4$ at the point $(-1,-1,2)$?

Let us regard the given surface as a particular level surface of the function $\phi = xy^3z^2$. Then the gradient of this function at the point $(-1,-1,2)$ will be perpendicular to the level surface through $(-1,-1,2)$, which is the given surface. When the gradient has been found, the unit normal can be obtained at once by dividing the gradient by its magnitude.

$$\nabla\phi = y^3z^2\mathbf{i} + 3xy^2z^2\mathbf{j} + 2xy^3z\mathbf{k}\,\bigg|_{-1,-1,2} = -4\mathbf{i} - 12\mathbf{j} + 4\mathbf{k}$$

$$|\nabla\phi| = \sqrt{16 + 144 + 16} = 4\sqrt{11}$$

$$\frac{\nabla\phi}{|\nabla\phi|} = \frac{-4\mathbf{i} - 12\mathbf{j} + 4\mathbf{k}}{4\sqrt{11}} = -\frac{1}{\sqrt{11}}\mathbf{i} - \frac{3}{\sqrt{11}}\mathbf{j} + \frac{1}{\sqrt{11}}\mathbf{k}$$

It may be necessary to reverse the direction of this result by multiplying it by -1, depending on which side of the surface we wish the normal to extend.

The vector character of the operator ∇ suggests that we also consider dot and cross products in which it appears as one factor.

If $\mathbf{F} = F_1\mathbf{i} + F_2\mathbf{j} + F_3\mathbf{k}$ is a vector whose components are functions of x, y, and z, this leads to the combinations

$$\begin{aligned}\nabla \cdot \mathbf{F} &= \left(\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\right) \cdot (F_1\mathbf{i} + F_2\mathbf{j} + F_3\mathbf{k}) \\ &= \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}\end{aligned} \tag{11}$$

which is known as the **divergence** of the vector **F**, and

$$\begin{aligned}\nabla \times \mathbf{F} &= \left(\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\right) \times (F_1\mathbf{i} + F_2\mathbf{j} + F_3\mathbf{k}) \\ &= \mathbf{i}\left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}\right) - \mathbf{j}\left(\frac{\partial F_3}{\partial x} - \frac{\partial F_1}{\partial z}\right) + \mathbf{k}\left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right) \\ &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ F_1 & F_2 & F_3 \end{vmatrix}\end{aligned} \tag{12}$$

which is known as the **curl** of **F**

Both the divergence and the curl admit of physical interpretations which justify their names For instance, to illustrate the significance of the divergence, consider a region of space filled with a moving fluid, and let

$$\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$$

be a vector function representing at each point the velocity with which the particle of fluid instantaneously at that point is moving If we fix our attention on an infinitesimal volume (Fig 12 14) in the region occupied by the fluid, there will be flow through each of its faces, and as a result the amount of fluid within the element may vary To measure this variation, let us compute the loss of fluid from the element in the time Δt

Now, the volume of fluid which passes through one face of the element ΔV in time Δt is approximately equal to the component of the fluid velocity normal to the face times the area of the face times Δt, and the corresponding mass flow is, of course, the product of this volume and the density of the fluid ρ Hence, computing the loss of fluid through each face in turn, remembering that since the fluid is not assumed to be incompressible the density

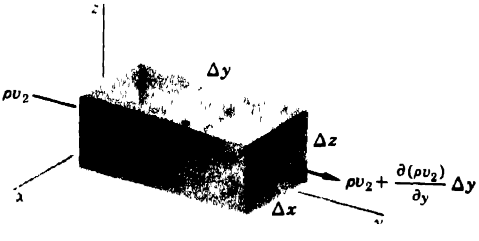


FIGURE 12.14
A typical volume element in a region filled with a moving fluid

as well as the velocity may vary from point to point, we have

$$
\begin{array}{ll}
\text{Right face:} & \left[\rho v_2 + \dfrac{\partial(\rho v_2)}{\partial y}\,\Delta y\right] \Delta x\,\Delta z\,\Delta t \\
\text{Left face:} & \qquad\qquad -\rho v_2\,\Delta x\,\Delta z\,\Delta t \\
\text{Front face} & \left[\rho v_1 + \dfrac{\partial(\rho v_1)}{\partial x}\,\Delta x\right] \Delta y\,\Delta z\,\Delta t \\
\text{Rear face:} & \qquad\qquad -\rho v_1\,\Delta y\,\Delta z\,\Delta t \\
\text{Top face:} & \left[\rho v_3 + \dfrac{\partial(\rho v_3)}{\partial z}\,\Delta z\right] \Delta x\,\Delta y\,\Delta t \\
\text{Bottom face} & \qquad\qquad -\rho v_3\,\Delta x\,\Delta y\,\Delta t
\end{array}
$$

If we add these and convert the resulting estimate of the absolute loss of fluid from ΔV in the interval Δt into the loss per unit volume per unit time by dividing by $\Delta V\,\Delta t \equiv \Delta x\,\Delta y\,\Delta z\,\Delta t$, we obtain in the limit

$$\text{Rate of loss per unit volume} = \frac{\partial(\rho v_1)}{\partial x} + \frac{\partial(\rho v_2)}{\partial y} + \frac{\partial(\rho v_3)}{\partial z}$$

which is precisely the divergence of the vector $\rho\mathbf{v}$. Thus fluid mechanics affords one possible interpretation of the divergence as the rate of loss of fluid per unit volume.

If the fluid is incompressible, there can be neither gain nor loss of fluid in a general element. Hence, since the density ρ is constant for an incompressible fluid, we must have

$$\nabla \cdot (\rho\mathbf{v}) = \rho\nabla \cdot \mathbf{v} = 0 \qquad \text{or} \qquad \nabla \cdot \mathbf{v} = 0$$

which is known as the **equation of continuity** for incompressible fluids However, if ΔV encloses a source of fluid, then there is a net loss of fluid through the surface of ΔV equal to the amount *diverging* from the source Similar results, of course, hold for such things as electric and magnetic flux, which exhibit many of the properties of incompressible fluids

To find a possible interpretation of the curl, let us consider a body rotating with uniform angular speed ω about an axis l. Let us define the **vector angular velocity $\mathbf{\Omega}$** to be a vector of length ω, extending along l in the direction in which a right-handed screw would advance if subject to the same rotation as the body Finally, let $\mathbf{R}$ be the vector drawn from any point O on the axis l to an arbitrary point P in the body.

From Fig 12 15 it is evident that the radius at which P rotates is $|\mathbf{R}|\ \ |\sin\theta|$. Hence, the linear speed of P is

$$|\mathbf{v}| = \omega|\mathbf{R}|\ \ |\sin\theta| = |\mathbf{\Omega}|\ |\mathbf{R}| \cdot |\sin\theta| = |\mathbf{\Omega} \times \mathbf{R}|$$

Moreover, the vector velocity $\mathbf{v}$ is directed perpendicular to the plane of $\mathbf{\Omega}$ and $\mathbf{R}$, so that $\mathbf{\Omega}$, $\mathbf{R}$, and $\mathbf{v}$ form a right-handed system. Hence, the cross product $\mathbf{\Omega} \times \mathbf{R}$ gives not only the magnitude of $\mathbf{v}$ but the direction as well.

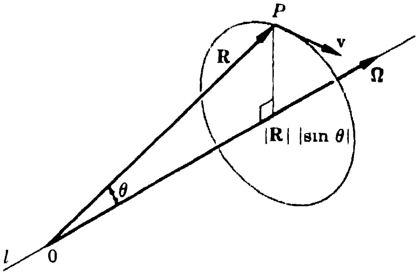


FIGURE 12.15
A physical interpretation of the curl.

Now, if we take the point O as the origin of coordinates, we can write

$$\mathbf{R} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \qquad \text{and} \qquad \mathbf{\Omega} = \Omega_1\mathbf{i} + \Omega_2\mathbf{j} + \Omega_3\mathbf{k}$$

Hence, the equation $\mathbf{v} = \mathbf{\Omega} \times \mathbf{R}$ can be written at length in the form

$$\mathbf{v} = (\Omega_2 z - \Omega_3 y)\mathbf{i} - (\Omega_1 z - \Omega_3 x)\mathbf{j} + (\Omega_1 y - \Omega_2 x)\mathbf{k}$$

If we take the curl of $\mathbf{v}$, we therefore have

$$\nabla \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ \Omega_2 z - \Omega_3 y & -(\Omega_1 z - \Omega_3 x) & \Omega_1 y - \Omega_2 x \end{vmatrix}$$

Expanding this, remembering that $\mathbf{\Omega}$ is a constant vector, we find

$$\nabla \times \mathbf{v} = 2\Omega_1\mathbf{i} + 2\Omega_2\mathbf{j} + 2\Omega_3\mathbf{k} = 2\mathbf{\Omega}$$

or $\qquad \mathbf{\Omega} = \tfrac{1}{2}\nabla \times \mathbf{v}$

The angular velocity of a uniformly rotating body is thus equal to one-half the curl of the linear velocity of any point of the body. The aptness of the name *curl* in this connection is apparent.

The results of applying the operator ∇ to various combinations of scalar and vector functions can be found by the following formulas.*

(13) $$\nabla \cdot (\phi\mathbf{v}) = \phi\nabla \cdot \mathbf{v} + \mathbf{v} \cdot \nabla\phi$$

(14) $$\nabla \times (\phi\mathbf{v}) = \phi\nabla \times \mathbf{v} + (\nabla\phi) \times \mathbf{v}$$

(15) $$\nabla \cdot (\mathbf{u} \times \mathbf{v}) = \mathbf{v} \cdot \nabla \times \mathbf{u} - \mathbf{u} \cdot \nabla \times \mathbf{v}$$

(16) $$\nabla \times (\mathbf{u} \times \mathbf{v}) = \mathbf{v} \cdot \nabla\mathbf{u} - \mathbf{u} \cdot \nabla\mathbf{v} + \mathbf{u}\nabla \cdot \mathbf{v} - \mathbf{v}\nabla \cdot \mathbf{u}$$

(17) $$\nabla(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot \nabla\mathbf{v} + \mathbf{v} \cdot \nabla\mathbf{u} + \mathbf{u} \times (\nabla \times \mathbf{v}) + \mathbf{v} \times (\nabla \times \mathbf{u})$$

(18) $$\nabla \times \nabla\phi = 0$$

(19) $$\nabla \cdot \nabla \times \mathbf{v} = 0$$

(20) $$\nabla \times (\nabla \times \mathbf{v}) = \nabla(\nabla \cdot \mathbf{v}) - \nabla \cdot \nabla\mathbf{v} = \nabla(\nabla \cdot \mathbf{v}) - \nabla^2\mathbf{v}$$

---

* We must remember, of course, that these results are correct only for the cartesian form of the operator ∇ given by Eq. (6). Different formulas arise when ∇ is expressed in terms of more general coordinate systems.

These identities can all be verified by direct expansion For instance, to prove (13), we have

$$\begin{aligned}\nabla\cdot(\phi\mathbf{v}) &= \nabla\cdot[\phi(v_1\mathbf{i}+v_2\mathbf{j}+v_3\mathbf{k})]\\ &= \frac{\partial(\phi v_1)}{\partial x}+\frac{\partial(\phi v_2)}{\partial y}+\frac{\partial(\phi v_3)}{\partial z}\\ &= \phi\frac{\partial v_1}{\partial x}+v_1\frac{\partial\phi}{\partial x}+\phi\frac{\partial v_2}{\partial y}+v_2\frac{\partial\phi}{\partial y}+\phi\frac{\partial v_3}{\partial z}+v_3\frac{\partial\phi}{\partial z}\end{aligned}$$

which, on regrouping, is simply

$$\phi\nabla\cdot\mathbf{v}+\mathbf{v}\cdot\nabla\phi \qquad \text{as asserted}$$

In general, however, it is easier to establish formulas like those in the above list by treating ∇ as a vector, manipulating the expressions according to the appropriate formulas from vector algebra, and finally giving ∇ its operational meaning. Since ∇ is a linear combination of scalar differential operators which obey the usual product rule of differentiation (that is, act on the factors in a product one at a time) it is clear that ∇ itself has this property In other words, we can apply ∇ to products of various sorts by assuming that each of the factors in turn is the only one which is variable and then adding the partial results so obtained As a notation to aid us in determining these partial results, it is helpful to attach to ∇, whenever it is followed by more than one factor, a subscript indicating the one factor upon which it is currently allowed to operate

To prove (14), using the second, more formal procedure, we suppose first that the scalar function ϕ is a constant, that is, we let ∇ operate only on the vector **v** Then we can write

$$\nabla_v\times(\phi\mathbf{v}) = \phi\nabla\times\mathbf{v}$$

where the subscript v has been omitted from the right-hand side, since it is always completely clear what ∇ operates on when it is followed by just one factor Similarly, if we regard **v** as constant and ϕ as variable, we have

$$\nabla_\phi\times(\phi\mathbf{v}) = (\nabla\phi)\times\mathbf{v}$$

where the parentheses now restrict the effect of ∇ to the factor ϕ alone and so make a subscript on ∇ unnecessary Finally, adding our two partial results, we have

$$\nabla_v\times(\phi\mathbf{v})+\nabla_\phi\times(\phi\mathbf{v}) \equiv \nabla\times(\phi\mathbf{v}) = \phi\nabla\times\mathbf{v}+(\nabla\phi)\times\mathbf{v}$$

To prove (15), we have, from the cyclic properties of scalar triple products,

$$\nabla_u\cdot(\mathbf{u}\times\mathbf{v}) = \mathbf{v}\cdot\nabla\times\mathbf{u} \qquad\text{and}\qquad \nabla_v\cdot(\mathbf{u}\times\mathbf{v}) = -\mathbf{u}\cdot\nabla\times\mathbf{v}$$

Hence, adding these two partial results, we find

$$\nabla\cdot(\mathbf{u}\times\mathbf{v}) = \mathbf{v}\cdot\nabla\times\mathbf{u}-\mathbf{u}\cdot\nabla\times\mathbf{v}$$

To prove (16), we have

$$\nabla_u \times (\mathbf{u} \times \mathbf{v}) = (\nabla_u \cdot \mathbf{v})\mathbf{u} - (\nabla_u \cdot \mathbf{u})\mathbf{v} = \mathbf{v} \cdot \nabla\mathbf{u} - \mathbf{v}\nabla \cdot \mathbf{u}$$

and $$\nabla_v \times (\mathbf{u} \times \mathbf{v}) = (\nabla_v \cdot \mathbf{v})\mathbf{u} - (\nabla_v \cdot \mathbf{u})\mathbf{v} = \mathbf{u}\nabla \cdot \mathbf{v} - \mathbf{u} \cdot \nabla\mathbf{v}$$

Hence, adding,

$$\begin{aligned}\nabla_u \times (\mathbf{u} \times \mathbf{v}) + \nabla_v \times (\mathbf{u} \times \mathbf{v}) &\equiv \nabla \times (\mathbf{u} \times \mathbf{v}) \\ &= \mathbf{v} \cdot \nabla\mathbf{u} - \mathbf{u} \cdot \nabla\mathbf{v} + \mathbf{u}\nabla \cdot \mathbf{v} - \mathbf{v}\nabla \cdot \mathbf{u}\end{aligned}$$

To prove (17) we note that

$$\begin{aligned}\mathbf{u} \times (\nabla \times \mathbf{v}) \equiv \mathbf{u} \times (\nabla_v \times \mathbf{v}) &= (\mathbf{u} \cdot \mathbf{v})\nabla_v - (\mathbf{u} \cdot \nabla)\mathbf{v} \\ &= \nabla_v(\mathbf{u} \cdot \mathbf{v}) - \mathbf{u} \cdot \nabla\mathbf{v}\end{aligned}$$

and $$\begin{aligned}\mathbf{v} \times (\nabla \times \mathbf{u}) \equiv \mathbf{v} \times (\nabla_u \times \mathbf{u}) &= (\mathbf{v} \cdot \mathbf{u})\nabla_u - (\mathbf{v} \cdot \nabla)\mathbf{u} \\ &= \nabla_u(\mathbf{u} \cdot \mathbf{v}) - \mathbf{v} \cdot \nabla\mathbf{u}\end{aligned}$$

Hence, transposing and adding, we find

$$\begin{aligned}\nabla_u(\mathbf{u} \cdot \mathbf{v}) + \nabla_v(\mathbf{u} \cdot \mathbf{v}) &\equiv \nabla(\mathbf{u} \cdot \mathbf{v}) \\ &= \mathbf{u} \times (\nabla \times \mathbf{v}) + \mathbf{v} \times (\nabla \times \mathbf{u}) + \mathbf{u} \cdot \nabla\mathbf{v} + \mathbf{v} \cdot \nabla\mathbf{u}\end{aligned}$$

Without explicit expansion we infer that (18) is correct, since the operational coefficient of ϕ, namely, $\nabla \times \nabla$, is, in effect, a cross product of identical factors and hence zero Similarly, without expansion, we infer that (19) is correct, since $\nabla \cdot \nabla \times \mathbf{v}$ is a scalar triple product containing two identical factors and hence is zero

To establish (20), we merely apply the usual rule for expanding a vector triple product

$$\nabla \times (\nabla \times \mathbf{v}) = (\nabla \cdot \mathbf{v})\nabla - (\nabla \cdot \nabla)\mathbf{v} = \nabla(\nabla \cdot \mathbf{v}) - \nabla^2\mathbf{v}$$

where the conventional symbol ∇^2 has been substituted for the second-order operator

$$\begin{aligned}\nabla \cdot \nabla &= \left(\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\right) \cdot \left(\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\right) \\ &= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\end{aligned}$$

EXERCISES

In the following exercises $R = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, as usual, and $r = |\mathbf{R}| = \sqrt{x^2 + y^2 + z^2}$

1 Prove that $\nabla \times \mathbf{R} = 0$. What is $\nabla \cdot \mathbf{R}$?

2 If $\mathbf{A}$ is an arbitrary constant vector, prove that $\nabla(\mathbf{A} \cdot \mathbf{R}) = \mathbf{A}$ What is $(\mathbf{A} \cdot \nabla)\mathbf{R}$?

3 Compute (a) the divergence and (b) the curl of the vector

$$xyz\mathbf{i} + 3x^2y\mathbf{j} + (xz^2 - y^2z)\mathbf{k}$$

4 What is the directional derivative of the function $2xy + z^2$ at the point $(1,-1,3)$ in the direction of the vector $\mathbf{i} + 2\mathbf{j} + 2\mathbf{k}$?

5 What is the unit normal to the surface $z = x^2 + y^2$ at the point $(1,-2,5)$?

6 What is the angle between the normals to the surface $xy = z^2$ at the points $(1,4,-2)$ and $(-3,-3,3)$?

7 Verify Eqs. (14) and (15) by direct expansion.

8 What is the generalization of Eq. (10) to the case in which ϕ is a function of u, v, and w, where u, v, and w are each functions of x, y, and z?

9 Prove that the curl of any vector whose direction is constant is perpendicular to that direction.

10 Prove that $(\mathbf{A} \times \nabla) \times \mathbf{R} = -2\mathbf{A}$ What is $(\mathbf{A} \times \nabla) \cdot \mathbf{R}$?

11 Prove that $\nabla \cdot [(\mathbf{A} \times \mathbf{R})/r] = 0$ for any constant vector $\mathbf{A}$.

12 Prove that $\nabla \times \left(\dfrac{\mathbf{A} \times \mathbf{R}}{r}\right) = \dfrac{\mathbf{A}}{r} + \dfrac{\mathbf{A} \cdot \mathbf{R}}{r^3}\mathbf{R}$, for any constant vector $\mathbf{A}$

13 Prove that $\nabla r^n = nr^{n-2}\mathbf{R}$

14 For what values of n is $\nabla^2 r^n = 0$?

15 Determine n so that $\nabla \cdot (r^n\mathbf{R})$ will vanish identically

16 Prove that the curl of $f(r)\mathbf{R}$ is identically zero

17 Prove that $\nabla\phi_1 \times \nabla\phi_2 = \nabla \times (\phi_1 \nabla\phi_2) = -\nabla \times (\phi_2 \nabla\phi_1)$.

18 If $u = x + y + z$, $v = x + y$, and $w = -2xz - 2yz - z^2$, show that $[\nabla u \; \nabla v \; \nabla w] = 0$.

19 If three functions u, v, and w are connected by a relation $f(u,v,w) = 0$, prove that $[\nabla u \; \nabla v \; \nabla w] = 0$. (Hint Consider the dot product of ∇f and $\nabla u \times \nabla v$)

20 If $\mathbf{V}_1$ and $\mathbf{V}_2$ are the vectors which join the fixed points $P_1(x_1,y_1,z_1)$ and $P_2.(x_2,y_2,z_2)$ to the variable point $P:(x,y\,z)$, prove that the gradient of $\mathbf{V}_1 \cdot \mathbf{V}_2$ is $\mathbf{V}_1 + \mathbf{V}_2$ What is $\nabla \cdot (\mathbf{V}_1 \times \mathbf{V}_2)$? What is $\nabla \times (\mathbf{V}_1 \times \mathbf{V}_2)$?

12.4

Line, surface, and volume integrals

In the rest of our work in vector analysis and in much of the work ahead of us in the chapters on complex variables, a simple extension of the familiar process of integration known as *line integration* will be of fundamental importance Although in vector analysis we are usually concerned with line integrals taken along space curves, it is convenient to begin our discussion with a consideration of line integration along plane curves, since the applications of line integration in our study of complex variables will be exclusively in two dimensions In both the two-dimensional and the three-dimensional case, our work will involve only continuous curves which are **sectionally smooth**; that is, curves which are continuous and consist of a finite number of arcs on each of which the tangent changes continuously Clearly, such curves can have at most a finite number of "corners" where the direction of the tangent changes abruptly Moreover, as we learned in calculus, the length of such a curve between any two of its points is finite

Let $F(x,y)$ be a function of x and y, and let C be a continuous, sectionally smooth curve joining the points A and B.† Furthermore, let the arc of C between A and B be divided into n segments Δs_i whose projections on the x- and y-axes are, respectively, Δx_i and Δy_i, and let (ξ_i,η_i) be the coordinates of an arbitrary point in the segment Δs_i (Fig. 12.16)

If we evaluate the given function $F(x,y)$ at each of the points

---

† $F(x,y)$ bears no relation to the equation of C and is merely a function defined at every point of the portion of the curve C under consideration.

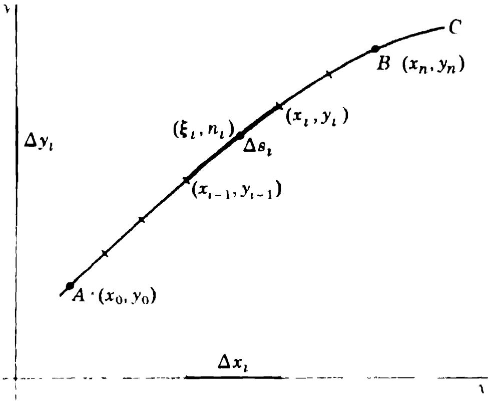


FIGURE 12.16 The subdivision of an arc preparatory to the definition of a line integral

(ξ_i,η_i) and form the products

$$F(\xi_i,\eta_i)\,\Delta x_i \qquad F(\xi_i,\eta_i)\,\Delta y_i \qquad F(\xi_i,\eta_i)\,\Delta s_i$$

and then sum over all the subdivisions of the arc AB, we have the three sums

$$\sum_{i=1}^{n} F(\xi_i,\eta_i)\,\Delta x_i \qquad \sum_{i=1}^{n} F(\xi_i,\eta_i)\,\Delta y_i \qquad \sum_{i=1}^{n} F(\xi_i,\eta_i)\,\Delta s_i$$

The limits of these sums, as n becomes infinite in such a way that the length of each Δs_i approaches zero, are known as **line integrals** and are written, respectively,

$$\int_C F(x,y)\,dx \qquad \int_C F(x,y)\,dy \qquad \int_C F(x,y)\,ds$$

It can be shown* that the continuity of $F(x,y)$ is a sufficient condition for the existence of the limits which define these integrals

In these definitions, Δx_i and Δy_i are signed quantities, whereas Δs_i is intrinsically positive Thus the following properties of ordinary definite integrals

a $$\int_A^B c\phi(t)\,dt = c\int_A^B \phi(t)\,dt \qquad c \text{ a constant}$$

b $$\int_A^B [\phi_1(t) \pm \phi_2(t)]\,dt = \int_A^B \phi_1(t)\,dt \pm \int_A^B \phi_2(t)\,dt$$

c $$\int_A^B \phi(t)\,dt = -\int_B^A \phi(t)\,dt$$

d $$\int_A^P \phi(t)\,dt + \int_P^B \phi(t)\,dt = \int_A^B \phi(t)\,dt$$

are equally valid for line integrals of the first two types, provided that throughout each formula the curve joining A and B remains the same. On the other hand, line integrals of the third type, although they do have properties **a** and **b**, do not have property **c**, since, in fact,

---

* See, for instance, D V Widder, "Advanced Calculus," p. 187, Prentice-Hall, Inc , Englewood Cliffs, N.J., 1947

$$\int_A^B F(x,y)\,ds = \int_B^A F(x,y)\,ds$$

Moreover, property **d** holds for these integrals if and only if P is between A and B on the path of integration In general, we shall be much more interested in integrals of the first two types than in integrals of the third type

Much of the initial strangeness of line integrals will disappear if we observe that the ordinary definite integrals of elementary calculus are just line integrals in which the curve C is the x-axis and the integrand is a function of x alone. Moreover, the evaluation of line integrals can be reduced to the evaluation of ordinary definite integrals, as the following example shows

EXAMPLE 1

What is the value of $\int_A^B \frac{1}{x+y}\,dx$ along each of the three paths shown in Fig 12 17?

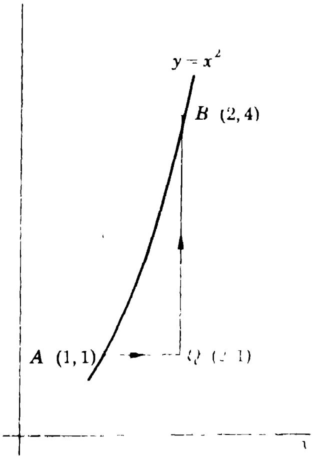


FIGURE 12.17 Three possible paths for line integration from A (1,1) to B (2,4)

Before this integral can be evaluated, it is necessary that y be expressed in terms of x To do this, we recall from the definition of a line integral that the integrand is always to be evaluated *along the path of integration* Along $y = x^2$ this gives us the ordinary definite integral

$$\int_1^2 \frac{dx}{x+x^2} = \int_1^2 \left(\frac{1}{x} - \frac{1}{1+x}\right) dx = [\ln x - \ln(1+x)]_1^2 = \ln\frac{4}{3}$$

Along AP the integral is obviously zero, since x remains constant Along PB, on which $y = 4$, we have the integral

$$\int_1^2 \frac{dx}{x+4} = [\ln(x+4)]_1^2 = \ln\frac{6}{5}$$

which is thus the value of the integral along the entire path APB. Along AQ, on which $y = 1$,

we have the integral

$$\int_1^2 \frac{dx}{x+1} = [\ln(x+1)]_1^2 = \ln\frac{3}{2}$$

Along QB the integral is again zero Hence, along the entire path AQB the value of the integral is ln $3/2$

This example not only illustrates the computational details of line integration, but also shows that in general a line integral depends not only on the end points of the integration but also upon the particular path which joins them

It is possible, as in the case of ordinary integration, to interpret a line integral as an area For, if we think of the integrand function $F(x,y)$ as defining a surface over the xy-plane, then the vertical cylindrical surface standing on the arc AB as base, or directrix, will cut the surface $z = F(x,y)$ in some curve such as PQ in Fig 12.18. This curve is clearly the upper boundary of the portion $ABQP$ of the cylindrical surface which lies above the xy-plane, below the surface $z = F(x,y)$, and between the generators AP and BQ Moreover, the product $F(\xi_i,\eta_i)\ \Delta s_i$ is approximately the area of the vertical strip of this portion of the surface which stands above the infinitesimal base Δs_i Hence, the sum

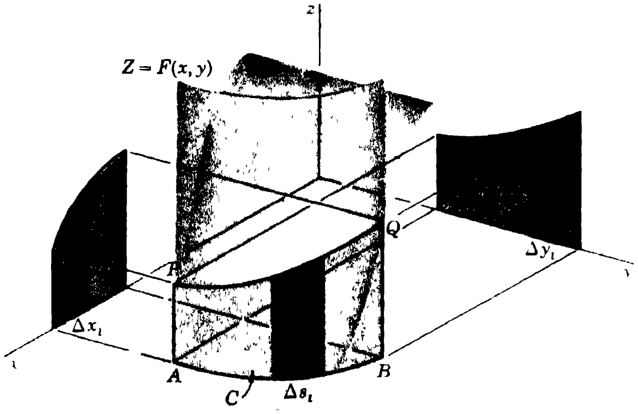


FIGURE 12.18
Plot showing the interpretation of a line integral as an area

$$\sum_{i=1}^{n} F(\xi_i,\eta_i)\ \Delta s_i$$

is approximately equal to the curved area $ABQP$, and, in the limit, the integral

$$\int_C F(x,y)\ ds$$

gives this area exactly.

In a similar fashion, the product $F(\xi_i,\eta_i)\ \Delta x_i$ is approximately the area of the projection on the xz-plane of the vertical strip standing on Δs_i; the sum

$$\sum_{i=1}^{n} F(\xi_i,\eta_i)\,\Delta x_i$$

represents approximately the area of the projection on the xz-plane of the entire curved area $ABQP$, and, in the limit, the integral

$$\int_C F(x,y)\,dx$$

gives the projected area exactly. In the same way the integral

$$\int_C F(x,y)\,dy$$

represents the area of the projection of $ABQP$ on the yz-plane.

Although this geometrical interpretation of line integrals as areas is vivid and easily grasped, it obscures the fact that almost invariably in applications the function $F(x,y)$ describes some physical property of the plane of integration and is actually unrelated to any other region of space

EXAMPLE 2

If a particle is attracted toward the origin by a force whose magnitude is proportional to the distance r of the particle from the origin, how much work is done when the particle is moved from the point $(0,1)$ to the point $(1,2)$ along the path $y = 1 + x^2$, assuming a coefficient of friction μ between the particle and the path?

Let θ be the angle which the tangent to the curve at a general point P (x,y) makes with the x-axis, let ϕ be the angle which the radius vector to P makes with the x-axis, and let α be the angle between the tangent and the radius vector at P (Fig 12 19) In moving the particle an infinitesimal distance Δs along the path, work must be done against two forces, namely, the tangential component of the central force

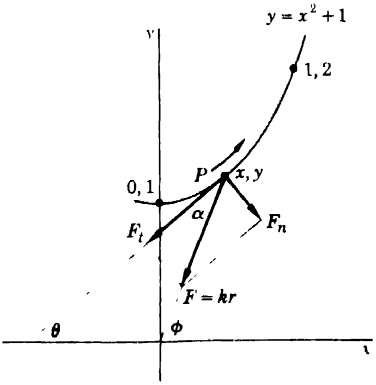
$$F_t = F\cos\alpha = kr\cos\alpha$$

and the frictional force

$$F_f = \mu F_n = \mu F\sin\alpha = \mu kr\sin\alpha$$

arising from the component of the central force which is normal to the path and which acts to press the particle against the path The infinitesimal amount of work done against these forces

FIGURE 12.19
The resolution of a central force into tangential and normal components along a curve



in moving a distance Δs is approximately

$$\Delta W = F_t\,\Delta s + F_f\,\Delta s = (kr\cos\alpha + \mu kr\sin\alpha)\,\Delta s$$

Now, from the exterior-angle theorem of plane geometry, $\alpha = \phi - \theta$ Hence,

$$\begin{aligned} W &= k\int_{0,1}^{1,2} r\cos(\phi-\theta)\,ds + \mu k\int_{0,1}^{1,2} r\sin(\phi-\theta)\,ds \\ &= k\int_{0,1}^{1,2} r(\cos\phi\cos\theta + \sin\phi\sin\theta)\,ds + \mu k\int_{0,1}^{1,2} r(\sin\phi\cos\theta - \cos\phi\sin\theta)\,ds \end{aligned}$$

But $\qquad r\cos\phi = x \qquad r\sin\phi = y$

and $\qquad \cos\theta\,ds = dx \qquad \sin\theta\,ds = dy$

Hence, substituting these into the last expression for W, we have

$$W = k\int_{0,1}^{1,2}(x\,dx + y\,dy) + \mu k\int_{0,1}^{1,2}(y\,dx - x\,dy)$$

The first of these integrals can be written very simply as

$$\frac{k}{2}\int_{0,1}^{1,2} d(x^2 + y^2)$$

which, *independent of the path*, is just

$$\frac{k}{2}(x^2 + y^2)\Big|_{0,1}^{1,2} = 2k$$

The second integral is not an exact differential, and thus, as usual, due account must be taken of the path Now, along the path, $y = x^2 + 1$ and $x = \sqrt{y-1}$ Hence,

$$\begin{aligned} \mu k\int_{0,1}^{1,2}(y\,dx - x\,dy) &= \mu k\int_0^1 (x^2+1)\,dx - \mu k\int_1^2 \sqrt{y-1}\,dy \\ &= \mu k\left[\frac{x^3}{3} + x\right]_0^1 - \mu k\left[\frac{2(y-1)^{3/2}}{3}\right]_1^2 = \frac{2\mu k}{3} \end{aligned}$$

The total amount of work done in the course of the motion is therefore

$$2k + \frac{2\mu k}{3}$$

The first term represents recoverable work stored as potential energy, the second term represents irrecoverable work dissipated as heat through friction

The extension of line integration to paths in three dimensions is easily accomplished Let $F(x,y,z)$ be a continuous function of x, y, and z, and let C be a continuous, sectionally smooth curve joining the points A and B Furthermore, let the arc of C between A and B be divided in an arbitrary manner into n subintervals Δs_i whose projections on the coordinate axes are Δx_i, Δy_i, and Δz_i, and let an arbitrary point P_i (ξ_i,η_i,ζ_i) be chosen in each Δs_i We now evaluate $F(x,y,z)$ at each of the points P_i and form the sums

$$\sum_{i=1}^n F(\xi_i,\eta_i,\zeta_i)\,\Delta x_i \qquad \sum_{i=1}^n F(\xi_i,\eta_i,\zeta_i)\,\Delta y_i \qquad \sum_{i=1}^n F(\xi_i,\eta_i,\zeta_i)\,\Delta z_i \qquad \sum_{i=1}^n F(\xi_i,\eta_i,\zeta_i)\,\Delta s_i$$

The limits of these sums as n becomes infinite in such a way that the length of each Δs_i approaches zero define the respective line integrals:

$$\int_C F(x,y,z)\,dx \qquad \int_C F(x,y,z)\,dy \qquad \int_C F(x,y,z)\,dz \qquad \int_C F(x,y,z)\,ds$$

Because of the difficulty of defining a space curve C as the intersection of several surfaces, it is customary to use a parametric representation for C. Hence, line integrals in three dimensions are ordinarily evaluated by integrating in terms of the parameter on C after the variables in the integrand have been replaced by their expressions in terms of the parameter.

EXAMPLE 3

What is $\int_C (xy + z^2)\,ds$, where C is the arc of the helix

$$x = \cos t \qquad y = \sin t \qquad z = t$$

which joins the points $(1,0,0)$ and $(-1,0,\pi)$?

Since $\qquad (ds)^2 = (dx)^2 + (dy)^2 + (dz)^2$

and since $dx = -\sin t\,dt$, $dy = \cos t\,dt$, and $dz = dt$, we have at once

$$ds = \sqrt{\sin^2 t + \cos^2 t + 1}\,|dt| = \sqrt{2}\,|dt|$$

Furthermore, it is clear that the point $(1,0,0)$ corresponds to the parametric value $t = 0$ and that the point $(-1,0,\pi)$ corresponds to the parametric value $t = \pi$. Hence, expressing the integrand in terms of the parameter t, the required integral becomes

$$\int_0^\pi (\cos t \sin t + t^2)\sqrt{2}\,dt = \sqrt{2}\left[\frac{\cos^2 t}{2} + \frac{t^3}{3}\right]_0^\pi = \frac{\sqrt{2}\,\pi^3}{3}$$

The concept of a line integral generalizes at once to *surface* and *volume integrals*. To describe the former, let $F(x,y,z)$ be a continuous function of x, y, and z, and let S be a given regular* surface or portion of a regular surface in the region of definition of $F(x,y,z)$. Let S be subdivided in an arbitrary manner into n elements ΔS_i (Fig. 12.20), and in each element let an arbitrary point P_i (ξ_i,η_i,ζ_i) be chosen. Finally, let $F(x,y,z)$ be evaluated at each of the points P_i. Then the limit of the sum

$$\sum_{i=1}^{n} F(\xi_i,\eta_i,\zeta_i)\,\Delta S_i$$

---

* A surface is said to be **smooth** if at each of its points there exists a tangent plane which varies continuously as the point varies continuously on the surface. A smooth surface is said to be **orientable** if it is two-sided; that is, if it is possible at each point to identify consistently a unique direction normal to the surface. A surface which can be subdivided by a finite number of sectionally smooth curves into pieces each of which is orientable (and therefore smooth) is said to be **regular**. For a discussion of smooth surfaces which are not regular, that is, smooth one-sided surfaces, see, for instance, Richard Courant and Herbert Robbins, "What Is Mathematics?", pp. 259–264, Oxford Book Company, Inc., New York, 1951.

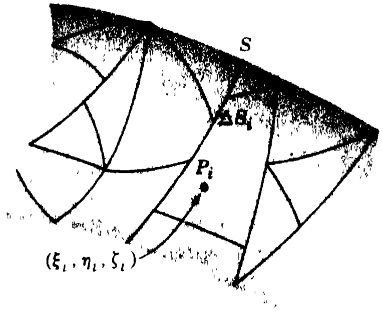


FIGURE 12.20 The subdivision of a surface preparatory to the definition of a surface integral.

as n becomes infinite in such a way that not only the area of each ΔS_i but also its maximum chord approaches zero, is the **surface integral**

$$\iint_S F(x,y,z)\,dS$$

Similarly, given a function $F(x,y,z)$ and a region of space V, we can subdivide V into arbitrary subregions ΔV_i, then evaluate $F(x,y,z)$ at an arbitrary point $P_i(\xi_i,\eta_i,\zeta_i)$ in each ΔV_i and form the sum

$$\sum_{i=1}^{n} F(\xi_i,\eta_i,\zeta_i)\,\Delta V_i$$

The limit of this sum as n becomes infinite in such a way that not only the volume of each ΔV_i but also its maximum chord approaches zero, is the **volume integral**

$$\iiint_V F(x,y,z)\,dV$$

EXAMPLE 4

What is the integral of the function x^2z over the entire surface of the right circular cylinder of height h which stands on the circle $x^2 + y^2 = a^2$? What is the integral of the given function throughout the volume of the cylinder?

To answer the first question, we must perform three integrations, i.e., we must integrate separately over the curved surface, the lower base, and the upper base of the cylinder. In each case, of course, we must employ a subdivision of the appropriate portion of the surface which will lead to integrals that can conveniently be evaluated. This is most easily done by using polar coordinates, as shown in Fig. 12.21. Then, on the curved surface, say S_1, we have

$$dS_1 = a\,d\theta\,dz \qquad x = a\cos\theta \qquad z = z$$

and the integral

$$\iint_{S_1} x^2z\,dS_1 = \int_0^h \int_0^{2\pi} (a\cos\theta)^2 z(a\,d\theta\,dz) = a^3 \int_0^h z\left[\frac{\theta}{2} + \frac{\sin 2\theta}{4}\right]_0^{2\pi} dz$$

$$= \pi a^3 \int_0^h z\,dz = \frac{\pi a^3h^2}{2}$$

On the lower base, say S_2, we have

$$dS_2 = r\,dr\,d\theta \qquad x = r\cos\theta \qquad z = 0$$

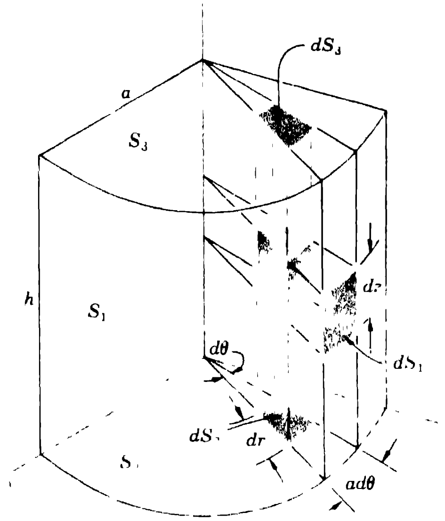


FIGURE 12.21
A typical volume element in cylindrical coordinates

However, because of the factor z, the integrand vanishes identically on S_2, and without further calculations we have

$$\iint_{S_2} x^2z\,dS_2 = 0$$

On the upper base, say S_3, we have $dS_3 = r\,dr\,d\theta$, $x = r\cos\theta$, and $z = h$ Hence,

$$\iint_{S_3} x^2z\,dS_3 = \int_0^{2\pi}\int_0^a (r\cos\theta)^2h(r\,dr\,d\theta) = h\int_0^{2\pi}\cos^2\theta\left[\frac{r^4}{4}\right]_0^a d\theta$$

$$= \frac{a^4h}{4}\left[\frac{\theta}{2}+\frac{\sin 2\theta}{4}\right]_0^{2\pi} = \frac{\pi a^4h}{4}$$

The integral over the entire surface S is, of course, the sum of the integrals over S_1, S_2, and S_3, i e ,

$$\iint_S x^2z\,dS = \frac{\pi a^4h^2}{2} + 0 + \frac{\pi a^4h}{4} = \frac{\pi a^4h(2h+a)}{4}$$

In computing the required volume integral it is also convenient to use polar coordinates Doing this, we have $dV = r\,dr\,d\theta\,dz$, $x = r\cos\theta$, $z = z$, and the integral

$$\iiint_V x^2z\,dV = \int_0^h\int_0^{2\pi}\int_0^a (r\cos\theta)^2z(r\,dr\,d\theta\,dz) = \frac{a^4}{4}\int_0^h\int_0^{2\pi} z\cos^2\theta\,d\theta\,dz$$

$$= \frac{\pi a^4}{4}\int_0^h z\,dz = \frac{\pi a^4h^2}{8}$$

For the most part, our interest in line, surface, and volume integrals will be theoretical rather than computational, that is, we shall use them far more often in derivations than in numerical calculation. Fundamental among the theorems we will need for this purpose is **Green's lemma,*** which relates the line integral of a function taken around the boundary of a plane region to the surface integral of an associated function taken over the region itself:

---

* Named for the English mathematical physicist George Green (1793–1841).

THEOREM 1

If R is a plane region bounded by a finite number of simple closed curves* and if $U(x,y)$, $V(x,y)$, $\frac{\partial U}{\partial y}$, and $\frac{\partial V}{\partial x}$ are continuous at all points of R and its boundary C, then

$$\int_C U\,dx + V\,dy = \iint_R \left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right) dx\,dy$$

PROOF Let us first suppose that the boundary of R is a single simple closed curve C with the property that any line parallel to either of the coordinate axes cuts it in at most two points, and let us draw the horizontal and vertical lines which circumscribe C (Fig. 12 22). Then the arcs $P_4P_1P_2$ and $P_4P_3P_2$ define single-valued functions of x, which we shall call $f_1(x)$ and $f_2(x)$, respectively Similarly, the arcs $P_1P_4P_3$ and $P_1P_2P_3$ define single-valued functions of y, which we shall call $g_1(y)$ and $g_2(y)$, respectively. Now consider

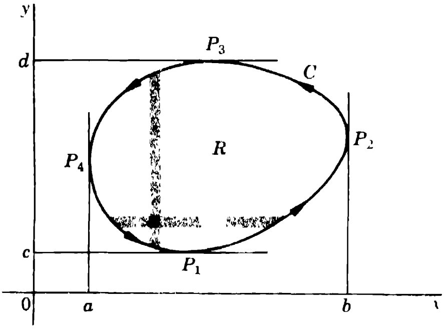


FIGURE 12.22
A plane region and its boundary

$$I_1 = \iint_R \frac{\partial V}{\partial x}\,dx\,dy$$

To carry out this integration over R, it is sufficient to integrate with respect to x from the arc $P_1P_4P_3$ to the arc $P_1P_2P_3$ and then to integrate with respect to y from c to d Hence,

$$I_1 = \int_c^d \int_{g_1(y)}^{g_2(y)} \frac{\partial V}{\partial x}\,dx\,dy$$

The inner integration can easily be performed, and we find

$$\begin{aligned} I_1 = \int_c^d V(x,y)\Big|_{g_1(y)}^{g_2(y)} dy &= \int_c^d V[g_2(y),y]\,dy - \int_c^d V[g_1(y),y]\,dy \\ &= \int_c^d V[g_2(y),y]\,dy + \int_d^c V[g_1(y),y]\,dy \end{aligned}$$

---

* For our purposes it is sufficient to define a simple closed curve as a closed, sectionally smooth curve which does not cross itself That this is not the whole story, however, can be inferred from the article "What Is a Curve?" by G T. Whyburn in the *Am Math Monthly*, vol 49, pp 493–497, October, 1942

Now, the first of these integrals is precisely the line integral

$$\int_c^d V(x,y)\,dy$$

taken along the path $x = g_2(y)$ from P_1 to P_3, and the second is just the same line integral taken along the path $x = g_1(y)$ *in the direction from* P_3*, through* P_4*, to* P_1. Together, then, they constitute the line integral of $V(x,y)$ around the entire closed curve C, hence,

$$(1)\qquad \iint_R \frac{\partial V}{\partial x}\,dx\,dy = \int_C V(x,y)\,dy$$

Similarly, if we consider

$$I_2 = \iint_R \frac{\partial U}{\partial y}\,dx\,dy = \iint_R \frac{\partial U}{\partial y}\,dy\,dx$$

we can write more specifically

$$I_2 = \int_a^b \int_{f_1(x)}^{f_2(x)} \frac{\partial U}{\partial y}\,dy\,dx$$

Performing the inner integration, we have

$$\begin{aligned} I_2 = \int_a^b U(x,y)\Big|_{f_1(x)}^{f_2(x)} dx &= \int_a^b U[x,f_2(x)]\,dx - \int_a^b U[x,f_1(x)]\,dx \\ &= -\int_b^a U[x,f_2(x)]\,dx - \int_a^b U[x,f_1(x)]\,dx \end{aligned}$$

The first of these integrals is just the negative of the line integral of $U(x,y)$ along $y = f_2(x)$ in the direction from P_2 to P_4. The second is the negative of the integral of $U(x,y)$ along $y = f_1(x)$ from P_4 to P_2. Together they constitute the negative of the line integral of $U(x,y)$ entirely around C in the same direction in which we integrated in (1)

$$(2)\qquad \iint_R \frac{\partial U}{\partial y}\,dx\,dy = -\int_C U(x,y)\,dx$$

If we subtract (2) from (1) and combine the integrals on each side, we obtain

$$(3)\qquad \int_C U\,dx + V\,dy = \iint_R \left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right) dx\,dy$$

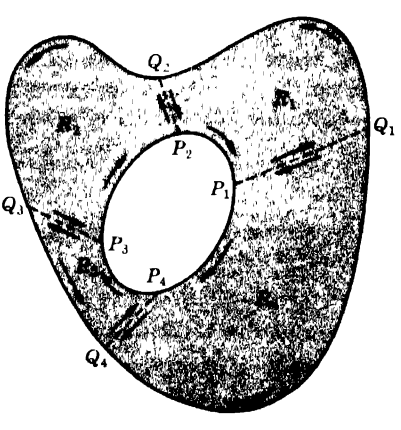


FIGURE 12.23 A plane region R subdivided into simpler regions R_1, R_2, R_3, R_4

which establishes Green's lemma for the special regions we have thus far been considering.

It is a simple matter, now, to extend Green's lemma to regions whose boundaries do not satisfy the condition that every line parallel to either of the coordinate axes cuts them in at most two points. For, if this is not the case, the region R can be divided into subregions R_i whose boundaries C_i do have this property. Then Eq. (3) can be applied to each subregion, following which the addition of these results yields Green's lemma for the general region R itself For instance, for the region shown in Fig. 12.23 we can subdivide as indicated and then apply Eq. (3) to each subregion, getting

$$\int_{C_1} U\,dx + V\,dy = \iint_{R_1}\left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right)dx\,dy$$

$$\int_{C_2} U\,dx + V\,dy = \iint_{R_2}\left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right)dx\,dy$$

$$\int_{C_3} U\,dx + V\,dy = \iint_{R_3}\left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right)dx\,dy$$

$$\int_{C_4} U\,dx + V\,dy = \iint_{R_4}\left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right)dx\,dy$$

When these results are added, the four integrals on the right combine to give exactly

$$\iint_{R}\left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right)dx\,dy$$

since $R_1 + R_2 + R_3 + R_4 = R$ Moreover, the four line integrals on the left combine to give the line integral around the two curves which form the boundary of R *plus* a set of line integrals taken along the auxiliary boundary arcs P_iQ_i Since U and V are continuous throughout R, these integrals cancel in pairs, however, since each of the segments P_iQ_i is traversed twice in opposite directions Hence we are left with

$$\int_{C} U\,dx + V\,dy = \iint_{R}\left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right)dx\,dy$$

which is the assertion of Green's lemma

The direction in which it is necessary to integrate around C, in order for Green's lemma to be correct as we have stated it, is characterized by the fact that an observer moving along C in this direction always has the interior of the region R on his left. This direction is called the **positive direction** of traversing C.

EXERCISES

1 Discuss the extension of Green's lemma to regions whose boundaries contain segments which are parallel to one or the other of the coordinate axes

2 Evaluate $\int_{0,1}^{2,3} (2xy - 1)\,dx + (x^2 + 1)\,dy$ along the paths $y = x + 1$ and $y = (x^2/2) + 1$

3 Evaluate $\int x^2y^2\,ds$ around the circle $x^2 + y^2 = 1$ (Hint: Use polar coordinates.)

4 Along what curve of the family $y = kx(1 - x)$ does the integral $\int_{0,0}^{1,0} y(x - y)\,dx$ attain its largest value?

5 Evaluate $\int_{-1,0}^{1,0} y(1 + x)\,dy$ (a) along the x-axis and (b) along $y = 1 - x^2$.

6 Evaluate $\int_{0,0}^{1,1} x\,ds$ along the paths $y = x$, $y = x^{1/2}$, and $y = x^2$.

7 Evaluate $\iint_S (x + y)z\,dS$, where S is the surface of the cube whose vertices are (0,0,0), (1,0,0), (1,1,0), (0,1,0), (0,0,1), (1,0,1), (1,1,1), and (0,1,1)

8 Evaluate $\iint_S (x + y + z)\,dS$, where S is the portion of the surface of the sphere $x^2 + y^2 + z^2 = a^2$ which lies in the first octant (Hint: Use spherical coordinates.)

9 Evaluate $\iiint_V x^2z\,dV$, where V is the volume under the surface $x^2 + y^2 + z^2 = a^2$ and above the xy-plane

10 Verify Green's lemma for the integral $\int(x^2 + y)\,dx - xy^2\,dy$, taken around the boundary of the square whose vertices are (0,0), (1,0), (1,1), and (0,1)

11 Verify Green's lemma for the integral $\int(x - y)\,dx + (x + y)\,dy$ taken around the boundary of the finite area in the first quadrant between the curves $y = x^2$ and $y^2 = x$.

12 Verify Green's lemma for the integral $\int(x - 2y)\,dx + x\,dy$ taken around the circle $x^2 + y^2 = a^2$

13 If a particle is attracted toward the origin by a force proportional to the nth power of the distance from the origin, show that the work done against this force in moving the particle from the point (x_0,y_0) to the point (x_1,y_1) is independent of the path, and find its amount.

14 A particle is attracted toward the origin by a force proportional to the cube of the distance from the origin How much work is done in moving the particle from the origin to the point (1,1) if motion takes place (a) along the path $y = x$, (b) along the path $y = x^2$, (c) along the x-axis to (1,0) and then vertically to (1,1), and (d) along the y-axis to (0,1) and then horizontally to (1,1), and if in each case the coefficient of friction between the particle and the path is μ?

15 If U, V, $\dfrac{\partial U}{\partial y}$, and $\dfrac{\partial V}{\partial x}$ are continuous and if $\dfrac{\partial U}{\partial y} = \dfrac{\partial V}{\partial x}$ at all points in the interior of simple closed curve C, show that $\int_\Gamma U\,dx + V\,dy = 0$ for any simple closed curve Γ which lies entirely within C

16 Show that Green's lemma fails to hold for the functions

$$U = -\frac{y}{x^2 + y^2} \qquad \text{and} \qquad V = \frac{x}{x^2 + y^2}$$

if R is the interior of the circle C: $x^2 + y^2 = 1$ Explain

17 Using Green's lemma, show that the area bounded by any simple closed curve C is given by the formula $A = \tfrac{1}{2}\int_C x\,dy - y\,dx$. Is this formula correct for regions bounded by more than one simple closed curve?

18 Using Green's lemma, establish the formula

$$\iint_R \left(\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2}\right) dx\,dy = \int_C \frac{dF}{dn}\,ds$$

where R is the region bounded by the simple closed curve C, and $\dfrac{\partial F}{\partial n}$ is the directional derivative of F in the direction of the outer normal to C

19 By setting $U = f\dfrac{\partial g}{\partial x}$ and $V = f\dfrac{\partial g}{\partial y}$ in Green's lemma, show that

$$\iint_R \left(\frac{\partial f}{\partial x}\frac{\partial g}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial g}{\partial x}\right) dx\,dy = \int_C f\,dg$$

where R is the region bounded by the simple closed curve C What is $\int_C g\,df$?

20 By setting $U = f\dfrac{\partial g}{\partial y}$ and $V = -f\dfrac{\partial g}{\partial x}$ in Green's lemma, show that

$$\iint_R f\left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}\right) dx\,dy + \iint_R \left(\frac{\partial f}{\partial x}\frac{\partial g}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial g}{\partial y}\right) dx\,dy = -\int_C f\frac{\partial g}{\partial n}\,ds$$

where R is the region bounded by the simple closed curve C, and $\dfrac{\partial g}{\partial n}$ is the directional derivative of g in the direction of the outer normal to C

12.5

Integral theorems

The integrals we encounter in vector analysis are in most cases scalar quantities For instance, given a vector function $\mathbf{F}(x,y,z)$, we are often interested in the integral of its tangential component along a curve C or in the integral of its normal component over a surface S In the first case, if $\mathbf{R}$ is the vector from the origin to a general point of C, so that $d\mathbf{R}/ds \equiv \mathbf{T}$ is the unit vector tangent to to C at a general point, then $\mathbf{F} \cdot \mathbf{T}$ is the tangential component of $\mathbf{F}$ and

$$\begin{aligned}\int_C \mathbf{F}\cdot\mathbf{T}\,ds &= \int_C \mathbf{F}\cdot\frac{d\mathbf{R}}{ds}\,ds \\ &= \int_C \mathbf{F}\cdot d\mathbf{R}\end{aligned} \tag{1}$$

is the integral of this component along the curve C In the second case, if $\mathbf{N}$ is the unit vector normal to S at a general point, then $\mathbf{F} \cdot \mathbf{N}$ is the normal component of $\mathbf{F}$ and

$$\iint_S \mathbf{F}\cdot\mathbf{N}\,dS\dagger \tag{2}$$

is the integral of this component over the surface S Other scalar integrals of frequent occurrence are the surface integral of the normal component of the curl of $\mathbf{F}$

$$\iint_S (\nabla \times \mathbf{F})\cdot\mathbf{N}\,dS \tag{3}$$

and the volume integral of the divergence of $\mathbf{F}$

$$\iiint_V \nabla\cdot\mathbf{F}\,dV \tag{4}$$

Fundamental in many of the applications of vector analysis is the so-called **divergence theorem,** which asserts the equality of the integrals (2) and (4) when V is the volume bounded by the closed regular surface S:

---

† Some writers denote the differential vector $\mathbf{N}\,dS$ by the symbol $d\mathbf{S}$ or $d\mathbf{A}$

THEOREM 1

If $\mathbf{F}(x,y,z)$ and $\nabla \cdot \mathbf{F}$ are continuous over the closed regular surface S and its interior V and if $\mathbf{N}$ is the unit vector perpendicular to S at a general point and extending outward from S, then

$$\iint_S \mathbf{N} \cdot \mathbf{F}\, dS = \iiint_V \nabla \cdot \mathbf{F}\, dV$$

PROOF To prove this theorem, we shall first suppose that S is a closed surface such that no line parallel to one of the coordinate axes cuts it in more than two points Now, if $\mathbf{F} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$, the assertion of the theorem can be written at length in the form

$$\iint_S \mathbf{N} \cdot (u\mathbf{i} + v\mathbf{j} + w\mathbf{k})\, dS = \iiint_V \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) dV$$

or

$$\iint_S \mathbf{N} \cdot \mathbf{i}u\, dS + \iint_S \mathbf{N} \cdot \mathbf{j}v\, dS + \iint_S \mathbf{N} \cdot \mathbf{k}w\, dS = \iiint_V \frac{\partial u}{\partial x}\, dV + \iiint_V \frac{\partial v}{\partial y}\, dV + \iiint_V \frac{\partial w}{\partial z}\, dV \tag{5}$$

We shall establish (5) by proving that respective integrals on each side are equal. To do this, let us consider first the integral

$$\iiint_V \frac{\partial w}{\partial z}\, dV$$

Under our assumption that no line parallel to one of the coordinate axes meets S in more than two points, it follows, in particular, that S is a double-valued surface over its projection on the xy-plane and, hence, can be thought of as consisting of a lower half, say S_1, and an upper half, say S_2 Then, if we take $dV = dx\, dy\, dz$ and perform the z-integration first, we have

$$\iiint_{z \text{ on } S_1}^{z \text{ on } S_2} \frac{\partial w}{\partial z}\, dz\, dx\, dy = \iint \left(w\Big|_{\text{on } S_2} - w\Big|_{\text{on } S_1} \right) dx\, dy \tag{6}$$

where, of course, x and y range over the area in the xy-plane which is the projection of S Moreover, the elements dS_1 and dS_2 can be defined so that they have $dx\, dy$ as their common projection on the xy-plane (Fig 12 24) Now $\mathbf{k} \cdot \mathbf{N}_1$ and $\mathbf{k} \cdot \mathbf{N}_2$

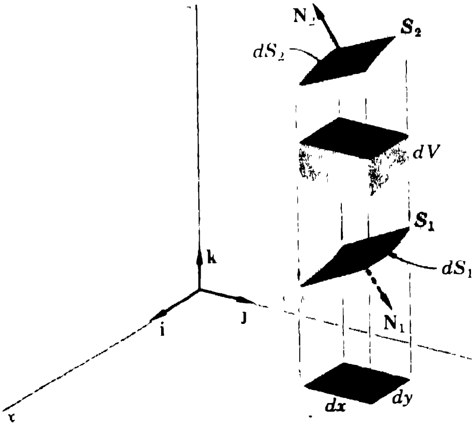


FIGURE 12.24 Integration in the z-direction from S_1 to S_2 in the proof of the divergence theorem

are, respectively, the cosines of the angles between the normal to the xy-plane $\mathbf{k}$ and the outer normals to dS_1 and dS_2; that is, they are numerically the cosines of the angles through which dS_1 and dS_2 are projected onto the element $dx\,dy$. Hence,

$$\begin{aligned} dx\,dy &= -\mathbf{k}\cdot\mathbf{N}_1\,dS_1 \\ &= \mathbf{k}\cdot\mathbf{N}_2\,dS_2 \end{aligned}$$

where the minus sign is necessary in the first equality because the outer normal $\mathbf{N}_1$ to dS_1 makes an angle of more than 90° with the direction of $\mathbf{k}$ and thus $\mathbf{k}\cdot\mathbf{N}_1$ is negative, whereas both $dx\,dy$ and dS_1 are clearly positive Therefore, substituting for $dx\,dy$ in the right-hand side of (6), that is, transferring the integration from the common projection of S_1 and S_2 back onto S_1 and S_2 themselves, we have

$$\begin{aligned} \iiint_V \frac{\partial w}{\partial z}\,dV &= \iint w\Big|_{\text{on } S_2} dx\,dy - \iint w\Big|_{\text{on } S_1} dx\,dy \\ &= \iint w\Big|_{\text{on } S_2} \mathbf{k}\cdot\mathbf{N}_2\,dS_2 + \iint w\Big|_{\text{on } S_1} \mathbf{k}\cdot\mathbf{N}_1\,dS_1 \\ &= \iint_{S_2} w\mathbf{k}\cdot\mathbf{N}\,dS + \iint_{S_1} w\mathbf{k}\cdot\mathbf{N}\,dS \end{aligned}$$

where the subscripts have been dropped from the integrands as superfluous, since the ranges of integration are now explicitly indicated. Finally, since S_1 and S_2 together make up the entire closed surface S, we can combine the last two integrals, getting

$$\iiint_V \frac{\partial w}{\partial z}\,dV = \iint_S w\mathbf{k}\cdot\mathbf{N}\,ds$$

Similarly we can show that

$$\iiint_V \frac{\partial u}{\partial x}\,dV = \iint_S u\mathbf{i}\cdot\mathbf{N}\,dS$$

$$\iiint_V \frac{\partial v}{\partial y}\,dV = \iint_S v\mathbf{j}\cdot\mathbf{N}\,dS$$

Adding the last three equations, we obtain the expanded form (5) of the divergence theorem, under the assumption that S is exactly two-valued over its projections on each of the coordinate planes

On the other hand, if S does not have this property, we can always subdivide its interior V into regions V_i whose boundaries S_i do have this property Then, applying our limited result to each of these regions, we obtain a set of equations of the form

$$\iint_{S_i} \mathbf{N}\cdot\mathbf{F}\,dS = \iiint_{V_i} \nabla\cdot\mathbf{F}\,dV$$

If these are added, the sum of the volume integrals is, of course, just the integral of $\nabla\cdot\mathbf{F}$ throughout the entire volume V The sum of the surface integrals is equal to the integral of $\mathbf{N}\cdot\mathbf{F}$ over the original surface S plus a set of integrals over the auxiliary boundary surfaces which were introduced when V was subdivided These cancel in pairs, however, since the integration extends twice over each interface, with integrands which are identical except for the oppositely directed unit normals they contain as factors Thus, our proof can be extended to volumes bounded by general closed regular surfaces, and Theorem 1 is established.

EXAMPLE 1

Prove that $\iint_S \mathbf{N} \times \mathbf{F}\, dS = \iiint_V \nabla \times \mathbf{F}\, dV$

To show this, let us apply the divergence theorem to the vector $\mathbf{F} \times \mathbf{C}$, where $\mathbf{C}$ is an arbitrary constant vector. Then

$$\iint_S \mathbf{N} \cdot (\mathbf{F} \times \mathbf{C})\, dS = \iiint_V \nabla \cdot (\mathbf{F} \times \mathbf{C})\, dV$$

Now taking advantage of the fact that $\mathbf{C}$ is a constant vector and that a cyclic permutation of the elements of a scalar triple product leaves the product unchanged, we can write

$$\iint_S \mathbf{C} \cdot \mathbf{N} \times \mathbf{F}\, dS = \iiint_V \mathbf{C} \cdot \nabla \times \mathbf{F}\, dV$$

or, removing the constant vector $\mathbf{C}$ from each integral,

$$\mathbf{C} \cdot \iint_S \mathbf{N} \times \mathbf{F}\, dS = \mathbf{C} \cdot \iiint_V \nabla \times \mathbf{F}\, dV$$

Since $\mathbf{C}$ is an arbitrary vector, this equation asserts that the vectors

$$\iint_S \mathbf{N} \times \mathbf{F}\, dS \qquad \text{and} \qquad \iiint_V \nabla \times \mathbf{F}\, dV$$

have equal projections in all directions and, hence, must be equal to each other, as asserted.

Various important theorems stem from the divergence theorem. For instance, if u and v are two sufficiently differentiable scalar point functions and if we set

$$\mathbf{F} = u\, \nabla v$$

then, by Eq. (13), Sec. 12.3,

$$\nabla \cdot \mathbf{F} = \nabla \cdot (u\, \nabla v) = u\nabla \cdot \nabla v + \nabla u \cdot \nabla v = \nabla u \cdot \nabla v + u\nabla^2 v$$

Hence, applying the divergence theorem to the vector $\mathbf{F} = u\, \nabla v$, we have

$$\iiint_V (\nabla u \cdot \nabla v + u\nabla^2 v)\, dV = \iint_S \mathbf{N} \cdot u\, \nabla v\, dS \tag{7}$$

Similarly, if we interchange the roles of u and v in (7), we obtain

$$\iiint_V (\nabla v \cdot \nabla u + v\nabla^2 u)\, dV = \iint_S \mathbf{N} \cdot v\, \nabla u\, dS \tag{8}$$

Finally, if we subtract (8) from (7), we obtain what is known as **Green's theorem**:*

THEOREM 2

If V is the volume bounded by a closed regular surface S and if $u(x,y,z)$ and $v(x,y,z)$ are scalar functions possessing continuous second partial derivatives, then

$$\iiint_V (u\nabla^2 v - v\nabla^2 u)\, dV = \iint_S \mathbf{N} \cdot (u\, \nabla v - v\, \nabla u)\, dS$$

Another result of some importance can be obtained by applying the divergence theorem to the function $\mathbf{F} = \mathbf{R}/r^3$, where, as usual,

$$\mathbf{R} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \qquad \text{and} \qquad r = |\mathbf{R}| = \sqrt{x^2 + y^2 + z^2}$$

---

* This should not be confused with *Green's lemma*, Theorem 1, Sec. 12.4.

Thus, substituting into the divergence theorem, we have

(9) $$\iint_S \mathbf{N}\cdot\frac{\mathbf{R}}{r^3}\,dS = \iiint_V \nabla\cdot\frac{\mathbf{R}}{r^3}\,dV$$

Now, by Eq. (13), Sec. 12.3, and Exercise 13, Sec. 12.3,

$$\nabla\cdot\frac{\mathbf{R}}{r^3} = \frac{1}{r^3}\nabla\cdot\mathbf{R} + \mathbf{R}\cdot\nabla\frac{1}{r^3} = \frac{3}{r^3} + \mathbf{R}\cdot\frac{d(1/r^3)}{dr}\nabla r$$
$$= \frac{3}{r^3} + \mathbf{R}\cdot\left(-\frac{3}{r^4}\cdot\frac{\mathbf{R}}{r}\right) = \frac{3}{r^3} - 3\frac{\mathbf{R}\cdot\mathbf{R}}{r^5} = 0$$

Hence, we conclude from (9) that

(10) $$\iint_S \mathbf{N}\cdot\frac{\mathbf{R}}{r^3}\,dS = 0$$

provided, of course, that r is different from zero at all points on and within S, that is, provided the origin from which **R** is drawn does not lie on S or within the volume enclosed by the surface S.

Since the divergence theorem requires that the function to which it is applied have continuous first partial derivatives throughout the volume of integration, it cannot be applied to $\mathbf{R}/r^3$ if the origin of **R** is within S. In this case we, therefore, modify the region of integration by constructing a sphere S' of radius ϵ having the origin O as center (Fig. 12.25). In the region V' between S and S' the function $\mathbf{R}/r^3$ satisfies the conditions of the divergence theorem, and thus Eq. (10) can properly be applied, giving

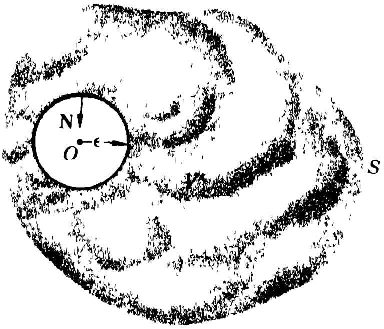


FIGURE 12.25 A singular point excluded from a three-dimensional region by an auxiliary spherical boundary.

(11) $$\iint_{S+S'} \mathbf{N}\cdot\frac{\mathbf{R}}{r^3}\,dS = 0 \qquad \text{or} \qquad \iint_S \mathbf{N}\cdot\frac{\mathbf{R}}{r^3}\,dS + \iint_{S'} \mathbf{N}\cdot\frac{\mathbf{R}}{r^3}\,dS = 0$$

Now, at any point of S', the direction of the normal which extends outward from the volume V' is opposite to **R**. Hence, the unit outer normal to S' is $\mathbf{N} = -\mathbf{R}/\epsilon$, since on S' the length of the radius vector **R** is $r = \epsilon$. Therefore, in the last integral,

$$\mathbf{N}\cdot\mathbf{R} = -\frac{\mathbf{R}}{\epsilon}\cdot\mathbf{R} = -\frac{\epsilon^2}{\epsilon} = -\epsilon$$

and Eq. (11) becomes

$$\iint_S \mathbf{N}\cdot\frac{\mathbf{R}}{r^3}\,dS + \iint_{S'} \frac{-\epsilon}{\epsilon^3}\,dS = 0$$

or
$$\iint_S \mathbf{N}\cdot\frac{\mathbf{R}}{r^3}\,dS = \frac{1}{\epsilon^2}\iint_{S'} dS = \frac{4\pi\epsilon^2}{\epsilon^2} = 4\pi$$

This result, coupled with Eq. (10), gives us **Gauss' theorem:**

THEOREM 3

If S is a closed regular surface, then

$$\iint_S \mathbf{N}\cdot\frac{\mathbf{R}}{r^3}\,dS = \begin{cases} 0 & O \text{ outside } S \\ 4\pi & O \text{ inside } S \end{cases}$$

Another integral formula of great importance in vector analysis is **Stokes' theorem:**[*]

THEOREM 4

If S is the portion of a regular surface bounded by the closed curve C and if $\mathbf{F}(x,y,z)$ is a vector function possessing continuous first partial derivatives, then

$$\int_C \mathbf{F}\cdot d\mathbf{R} = \iint_S \mathbf{N}\cdot\nabla\times\mathbf{F}\,dS$$

provided the direction of integration around C is positive with respect to the side of S on which the unit normals are drawn

PROOF To prove this, we suppose first that S has the property that it is single-valued above its projections on each of the coordinate planes. Now, if we write $\mathbf{F} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$, Stokes' theorem becomes

$$\begin{aligned} \int_C u\,dx + v\,dy + w\,dz &= \iint_S \mathbf{N}\cdot\nabla\times(u\mathbf{i} + v\mathbf{j} + w\mathbf{k})\,dS \\ &= \iint_S \mathbf{N}\cdot\nabla\times u\mathbf{i}\,dS + \iint_S \mathbf{N}\cdot\nabla\times v\mathbf{j}\,dS \\ &\quad + \iint_S \mathbf{N}\cdot\nabla\times w\mathbf{k}\,dS \end{aligned} \tag{12}$$

and, to establish it, it is sufficient to show that respective integrals on the two sides of the last equation are equal. We consider first the integral

$$\iint_S \mathbf{N}\cdot\nabla\times u\mathbf{i}\,dS$$

taken over the *closed* surface consisting of S, its projection on the xy-plane, say S', and the cylindrical surface, say S'', which projects S into S' (Fig. 12 26a) If we apply the divergence theorem to the vector $\nabla\times u\mathbf{i}$ over this surface and the volume it encloses, we obtain

$$\begin{aligned} \iint_S \mathbf{N}\cdot\nabla\times u\mathbf{i}\,dS + \iint_{S'} \mathbf{N}\cdot\nabla\times u\mathbf{i}\,dS + \iint_{S''} \mathbf{N}\cdot\nabla\times u\mathbf{i}\,dS \\ = \iiint_V \nabla\cdot(\nabla\times u\mathbf{i})\,dV = 0 \end{aligned}$$

or

$$\iint_S \mathbf{N}\cdot\nabla\times u\mathbf{i}\,dS = -\iint_{S'} \mathbf{N}\cdot\nabla\times u\mathbf{i}\,dS - \iint_{S''} \mathbf{N}\cdot\nabla\times u\mathbf{i}\,dS \tag{13}$$

---

[*] Named for the English mathematical physicist G G Stokes (1819–1903).

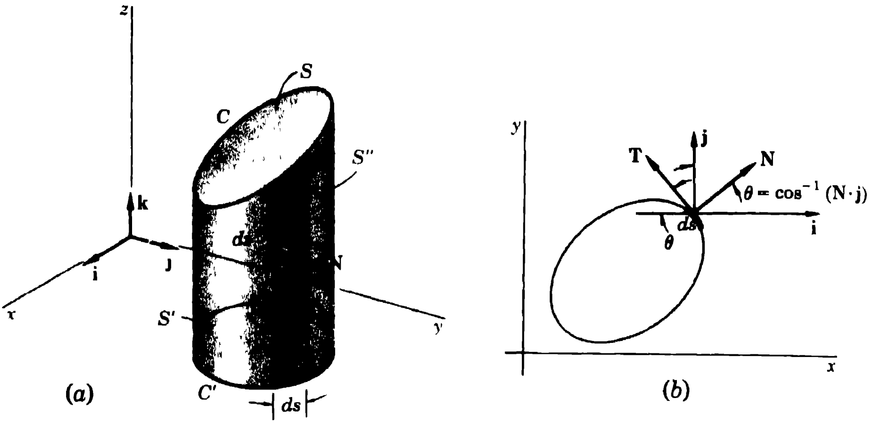


FIGURE 12.26
The closed surface $S + S' + S''$ employed in the proof of Stokes' theorem

since, as we showed in Sec 12 3, the divergence of the curl of any vector is identically zero. Now,

$$\nabla \times u\mathbf{i} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ u & 0 & 0 \end{vmatrix} = \mathbf{j}\frac{\partial u}{\partial z} - \mathbf{k}\frac{\partial u}{\partial y}$$

Moreover, on S' the outer normal $\mathbf{N}$ is clearly equal to $-\mathbf{k}$ Hence, on S' we have

$$\mathbf{N} \cdot \nabla \times u\mathbf{i} = -\mathbf{k} \cdot \left(\mathbf{j}\frac{\partial u}{\partial z} - \mathbf{k}\frac{\partial u}{\partial y}\right) = \frac{\partial u}{\partial y}$$

and
$$\iint_{S'} \mathbf{N} \cdot \nabla \times u\mathbf{i}\, dS = \iint_{S'} \frac{\partial u}{\partial y}\, dS$$

If we now apply Green's lemma (Theorem 1, Sec. 12 4) to the last integral, we find

$$\iint_{S'} \mathbf{N} \cdot \nabla \times u\mathbf{i}\, dS = -\int_{C'} u\, dx \tag{14}$$

Furthermore, since S'' is a cylindrical surface whose generators are parallel to the z-axis, the normals to S'' are all perpendicular to the vector $\mathbf{k}$ Therefore, on S'' we have

$$\mathbf{N} \cdot \nabla \times u\mathbf{i} = \mathbf{N} \cdot \left(\mathbf{j}\frac{\partial u}{\partial z} - \mathbf{k}\frac{\partial u}{\partial y}\right) = \mathbf{N} \cdot \mathbf{j}\frac{\partial u}{\partial z}$$

where, clearly, $\mathbf{N} \cdot \mathbf{j}$ is independent of z Then, taking $dS = dz\, ds$ (Fig 12 26a), we have

$$\begin{aligned} \iint_{S''} \mathbf{N} \cdot \nabla \times u\mathbf{i}\, dS &= \int_{C'} \int_{z \text{ on } S'}^{z \text{ on } S} \mathbf{N} \cdot \mathbf{j}\frac{\partial u}{\partial z}\, dz\, ds \\ &= \int_{C'} \left(u\Big|_S - u\Big|_{S'}\right) \mathbf{N} \cdot \mathbf{j}\, ds \end{aligned} \tag{15}$$

Now $\mathbf{N} \cdot \mathbf{j}$ is equal to the cosine of the angle between the normal $\mathbf{N}$ and the positive y-axis, and this is numerically equal but opposite in sign to the cosine of the angle between the directed tangent to C' and the positive x-axis (Fig. 12.26b).

Hence, $\mathbf{N}\cdot\mathbf{j}\,ds = -dx$, and Eq. (15) becomes

$$\iint_{S''} \mathbf{N}\cdot\nabla\times u\mathbf{i}\,dS = -\int_{C'} u\Big|_S dx + \int_{C'} u\Big|_{S'} dx \tag{16}$$

Now, in the first integral on the right in (16), the integrand, being evaluated at those points of S which are directly above the curve C', is actually evaluated along the curve C. Moreover, because C' is the projection of C in the z-direction, the variation of x around C' is exactly the same as the variation of x around C. Hence, in this integral we can properly replace the indicated path of integration C' by the curve C, getting

$$\iint_{S''} \mathbf{N}\cdot\nabla\times u\mathbf{i}\,dS = -\int_C u\,dx + \int_{C'} u\,dx \tag{17}$$

Therefore, substituting from (14) and (17) into (13), we have

$$\begin{aligned}\iint_S \mathbf{N}\cdot\nabla\times u\mathbf{i}\,dS &= -\left(-\int_{C'} u\,dx\right) - \left(-\int_C u\,dx + \int_{C'} u\,dx\right)\\ &= \int_C u\,dx\end{aligned} \tag{18}$$

In precisely the same way we can show that

$$\iint_S \mathbf{N}\cdot\nabla\times v\mathbf{j}\,dS = \int_C v\,dy \tag{19}$$

$$\iint_S \mathbf{N}\cdot\nabla\times w\mathbf{k}\,dS = \int_C w\,dz \tag{20}$$

Finally by adding (18), (19), and (20) we obtain Eq. (12).

It is now a simple matter to extend Eq. (12) to surfaces S which are not single-valued above their projections on the coordinate planes. For, if this is not the case, we can always subdivide S into regions S_i which do have this property and then apply Eq. (12) to each S_i and its boundary C_i, getting the set of equations

$$\int_{C_1} \mathbf{F}\cdot d\mathbf{R} = \iint_{S_1} \mathbf{N}\cdot\nabla\times\mathbf{F}\,dS$$

$$\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot$$

$$\int_{C_n} \mathbf{F}\cdot d\mathbf{R} = \iint_{S_n} \mathbf{N}\cdot\nabla\times\mathbf{F}\,dS$$

When these are added, the surface integrals combine to give precisely the surface integral over S itself, since $S_1 + \cdots + S_n = S$. At the same time the line integrals combine to give the line integral around the actual boundary of S plus the line integral along all the auxiliary boundary arcs taken twice in opposite directions (Fig. 12.27). Since the latter cancel identically, the line integral around C itself is all that remains, and the theorem follows in the general case.

If A and B are two arbitrary points in space, it is often important to know whether the line integral

$$\int_A^B \mathbf{F}\cdot d\mathbf{R} \tag{21}$$

is independent of the path which joins A and B. As a first step in establishing criteria for this, we observe that, if the integral (21) is independent of the path, then

$$\int \mathbf{F}\cdot d\mathbf{R}$$

taken around any closed path is zero. For let C be any simple

FIGURE 12.27 A portion of a surface S subdivided into simpler regions $S_1, S_2, \ldots, S_n$

closed curve, and let A and B be any two points on C (Fig. 12 28). Then, since the integral is independent of the path, by hypothesis, we have

$$\int_{\widehat{APB}} \mathbf{F} \cdot d\mathbf{R} = \int_{\widehat{AQB}} \mathbf{F} \cdot d\mathbf{R}$$

Now, if we reverse the direction of integration in the integral on the right, we have

$$\int_{\widehat{APB}} \mathbf{F} \cdot d\mathbf{R} = -\int_{\widehat{BQA}} \mathbf{F} \cdot d\mathbf{R}$$

or, transposing,

$$\int_{\widehat{APB}} \mathbf{F} \cdot d\mathbf{R} + \int_{\widehat{BQA}} \mathbf{F} \cdot d\mathbf{R} = \int_C \mathbf{F} \cdot d\mathbf{R} = 0 \qquad \text{as asserted}$$

Conversely, if $\int \mathbf{F} \cdot d\mathbf{R}$ is zero around every closed curve in a region, then the integral (21) is independent of the path For if $\widehat{APB}$ and $\widehat{AQB}$ are any two paths joining A and B (Fig 12 28), we have, by hypothesis,

$$\int_{\widehat{APB}} \mathbf{F} \cdot d\mathbf{R} + \int_{\widehat{BQA}} \mathbf{F} \cdot d\mathbf{R} = 0$$

whence, by reversing the direction of integration along $\widehat{BQA}$ and transposing,

$$\int_{\widehat{APB}} \mathbf{F} \cdot d\mathbf{R} = \int_{\widehat{AQB}} \mathbf{F} \cdot d\mathbf{R} \qquad \text{as asserted}$$

Now if the integral (21) is independent of the path, then when we integrate from a fixed point P_0 (x_0,y_0,z_0) to a variable

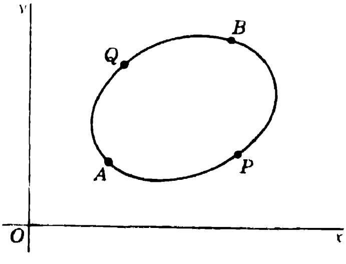


FIGURE 12.28 Two paths from A to B forming a simple closed curve.

point $P:(x,y,z)$, the result is a function only of the coordinates x, y, z of the variable end point. That is, if $\mathbf{F} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$, we can appropriately write

$$\int_{P_0}^{P} \mathbf{F} \cdot d\mathbf{R} = \int_{P_0}^{P} u\,dx + v\,dy + w\,dz = \phi(x,y,z)$$

In what follows it will be necessary to know the partial derivatives of the function ϕ defined by the last equation. To obtain these, it is convenient to go back to the fundamental definition of a derivative and write, for the x-partial derivative, for instance,

$$\begin{aligned}\frac{\partial\phi}{\partial x} &= \lim_{\Delta x\to 0} \frac{\phi(x+\Delta x, y, z) - \phi(x,y,z)}{\Delta x} \\ &= \lim_{\Delta x\to 0} \frac{1}{\Delta x}\left(\int_{x_0,y_0,z_0}^{x+\Delta x,y,z} u\,dx + v\,dy + w\,dz - \int_{x_0,y_0,z_0}^{x,y,z} u\,dx + v\,dy + w\,dz\right)\end{aligned}$$

Since by hypothesis these integrals are independent of the path, we can use any paths we find convenient In particular, in the integral from (x_0,y_0,z_0) to $(x + \Delta x, y, z)$, we shall let the path of integration consist of any smooth curve joining (x_0,y_0,z_0) to (x,y,z) plus the segment of the straight line joining (x,y,z) to $(x + \Delta x, y, z)$ (Fig. 12 29) Then,

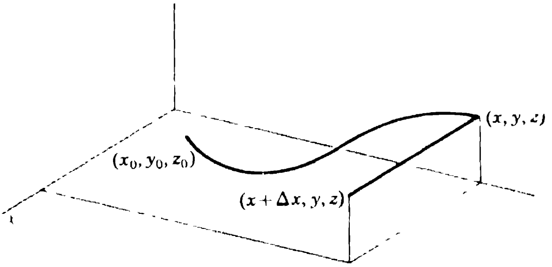


FIGURE 12.29
A convenient path of integration from (x_0,y_0,z_0) through (x,y,z) to $(x + \Delta x, y, z)$

$$\begin{aligned}\frac{\partial\phi}{\partial x} &= \lim_{\Delta x\to 0} \frac{1}{\Delta x}\left[\left(\int_{x_0,y_0,z_0}^{x,y,z} u\,dx + v\,dy + w\,dz + \int_{x,y,z}^{x+\Delta x,y,z} u\,dx + v\,dy + w\,dz\right)\right. \\ &\qquad\qquad \left. - \int_{x_0,y_0,z_0}^{x,y,z} u\,dx + v\,dy + w\,dz\right] \\ &= \lim_{\Delta x\to 0} \frac{1}{\Delta x}\int_{x,y,z}^{x+\Delta x,y,z} u\,dx + v\,dy + w\,dz\end{aligned}$$

Now, along the path of integration in the last integral, we have

$dy \equiv 0$ and $dz \equiv 0$

Hence, $$\frac{\partial\phi}{\partial x} = \lim_{\Delta x\to 0} \frac{1}{\Delta x}\int_{x}^{x+\Delta x} u\,dx$$

Since u is assumed to be continuous, the law of the mean for integrals can be applied to the last expression, and we have

$$\begin{aligned}\frac{\partial\phi}{\partial x} &= \lim_{\Delta x\to 0} \frac{1}{\Delta x}[u(x + \theta\,\Delta x, y, z)\,\Delta x] \qquad 0 < \theta < 1 \\ &= u(x,y,z)\end{aligned}$$

In the same way the partial derivatives with respect to y and z can be determined, and we have the following theorem:

THEOREM 5

If $\mathbf{F} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$ is a continuous function of x, y, and z with the property that

$$\int \mathbf{F} \cdot d\mathbf{R} = \int u\,dx + v\,dy + w\,dz$$

is independent of the path, then the partial derivatives of the function

$$\phi(x,y,z) \equiv \int_{P_0}^{P} \mathbf{F} \cdot d\mathbf{R} = \int_{P_0}^{P} u\,dx + v\,dy + w\,dz$$

are $$\frac{\partial \phi}{\partial x} = u \qquad \frac{\partial \phi}{\partial y} = v \qquad \frac{\partial \phi}{\partial z} = w$$

We are now in a position to show that, if $\mathbf{F} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$ is a continuous vector function and that, if $\int \mathbf{F} \cdot d\mathbf{R}$ is independent of the path, then $\mathbf{F}$ is the gradient of some scalar function ϕ In fact, if we define

$$\phi(x,y,z) = \int_{P_0}^{P} \mathbf{F} \cdot d\mathbf{R} = \int_{P_0}^{P} u\,dx + v\,dy + w\,dz$$

we have, by Theorem 5,

$$\nabla\phi \equiv \frac{\partial \phi}{\partial x}\mathbf{i} + \frac{\partial \phi}{\partial y}\mathbf{j} + \frac{\partial \phi}{\partial z}\mathbf{k} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k} = \mathbf{F}$$

as asserted

Before we can state a correct converse of the last result, we must distinguish between two types of regions in space On the one hand, a region V may have the property that every simple closed curve within it can be continuously contracted into a point without at any stage having to leave the region Regions of this type are called **simply connected**; as examples we have the interior of a sphere, the exterior of a sphere, and the space between two concentric spheres On the other hand, a region V may contain simple closed curves which cannot be continuously contracted into a point without at some stage having to leave the region. Such regions are called **multiply connected**; as an example we have the space between two infinitely long, coaxial cylinders, within which it is clearly impossible to shrink into a single point any closed curve encircling the inner cylindrical boundary Both the interior and the exterior of a torus are also examples of multiply connected regions.*

Now suppose that, throughout a simply connected region V, the vector function $\mathbf{F}$ is the gradient of a scalar function ϕ Then

$$\int_A^B \mathbf{F} \cdot d\mathbf{R} \equiv \int_A^B \nabla\phi \cdot dR = \int_A^B d\phi = \phi\Big|_A^B$$

and thus the integral of $\mathbf{F}$ depends only on the coordinates of the end points A and B and not on the path which joins them It is easy to show by an example (Exercise 30) that this is not necessarily true for multiply connected regions, since in such cases ϕ need not be continuous and single-valued throughout the region.

---

* The distinction between simply connected and multiply connected regions applies equally well in the plane, of course, and in our study of functions of a complex variable it will often be an important consideration

Finally, we observe that, if the curl of **F** is identically zero throughout a simply connected region V, then $\int \mathbf{F} \cdot d\mathbf{R}$ is independent of the path, and conversely For, if C is an arbitrary closed curve in a simply connected region V, it can be spanned by a surface S also lying entirely in V Then, by Stokes' theorem, we have

$$\int_C \mathbf{F} \cdot d\mathbf{R} = \iint_S \mathbf{N} \cdot \nabla \times \mathbf{F}\, dS$$

and, if $\nabla \times \mathbf{F} \equiv 0$, it follows that

$$\int_C \mathbf{F} \cdot d\mathbf{R} = 0$$

But, by one of our earlier observations, if $\int \mathbf{F} \cdot d\mathbf{R}$ is zero around every closed curve, then it is independent of the path, as asserted On the other hand, if $\int \mathbf{F} \cdot d\mathbf{R}$ is independent of the path, then, as we showed above, **F** is the gradient of a certain scalar function ϕ But then $\nabla \times \mathbf{F} = \nabla \times \nabla\phi$, and this is identically zero, by Eq. (18), Sec 12 3

The results of the preceding discussion can now be summarized in the following theorem

THEOREM 6

If $\mathbf{F} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$ is a function of x, y, and z possessing continuous first partial derivatives at all points of a simply connected region V, then the following statements are all equivalent, that is, any one of them implies each of the others

a $\int \mathbf{F} \cdot d\mathbf{R} \equiv \int u\,dx + v\,dy + w\,dz$ is independent of the path

b $\int \mathbf{F} \cdot d\mathbf{R} \equiv \int u\,dx + v\,dy + w\,dz$ is zero around every closed curve

c $\mathbf{F} \cdot d\mathbf{R} \equiv u\,dx + v\,dy + w\,dz$ is an exact differential

d **F** is the gradient of the scalar point function

$$\phi(x,y,z) = \int_{P_0}^{P} \mathbf{F} \cdot d\mathbf{R} \equiv \int_{P_0}^{P} u\,dx + v\,dy + w\,dz$$

e The curl of **F** vanishes identically

EXERCISES

1 If $\mathbf{F} = 2y\mathbf{i} + x\mathbf{j} + z^2\mathbf{k}$, evaluate $\int_{0,0,0}^{1,1,1} \mathbf{F}\ d\mathbf{R}$ along

a The rectilinear path from (0,0,0) to (1,0,0) to (1,1,0) to (1,1,1)

b The rectilinear path from (0,0,0) to (1,1,0) to (1,1,1)

c The straight line joining (0,0,0) to (1,1,1)

d The curve $x^2 + y^2 = 2z$, $x = y$

2 If $\mathbf{F} = x\mathbf{i} + y\mathbf{j} + 2\mathbf{k}$, evaluate $\iint_S \mathbf{F} \cdot \mathbf{N}\, dS$ over

a The surface of the cube whose vertices are (0,0,0), (1,0,0), (1,1,0), (0,1,0), (0,0,1), (1,0,1), (1,1,1), (0,1,1)

b The portion of the plane $x + 2y + 3z = 6$ which lies in the first octant c The entire surface of the sphere $x^2 + y^2 + z^2 = 1$

d The portion of the cone $x^2 + y^2 - (1 - z)^2 = 0$ between the planes $z = 0$ and $z = 1$

3 If $\mathbf{F} = y\mathbf{i} + x\mathbf{j} + z^2\mathbf{k}$, evaluate $\iiint_V \nabla \cdot \mathbf{F}\, dV$ throughout

a The volume bounded by the cube whose vertices are (0,0,0), (1,0,0), (1,1,0), (0,1,0), (0,0,1), (1,0,1), (1,1,1), (0,1,1)

b The volume cut off from the first octant by the plane $x + 2y + 3z = 6$

c The upper half of the volume within the sphere $x^2 + y^2 + z^2 = 1$

d The volume under the paraboloid $z = 1 - x^2 - y^2$ and above the plane $z = 0$

4 Write the divergence theorem in cartesian form

5 Write Green's theorem in cartesian form

6 Write Gauss' theorem in cartesian form

7 Write Stokes' theorem in cartesian form

8 If S is a closed surface, what is $\iint_S \mathbf{N} \cdot \nabla \times \mathbf{F}\, dS$?

9 If $\mathbf{T}$ is the variable unit tangent to a curve C, what is $\int_C \mathbf{T} \cdot d\mathbf{R}$? Can Stokes' theorem be used to evaluate this integral?

10 If $\mathbf{A}$ is a constant vector and C is a closed curve, show that $\int_C \mathbf{A} \cdot d\mathbf{R} = 0$ What is $\int_C d\mathbf{R}$?

11 If C is a closed curve, show that $\int_C \mathbf{R} \cdot d\mathbf{R} = 0$

12 If C is a closed curve, show that $\int_C (u\,\nabla v + v\,\nabla u) \cdot d\mathbf{R} = 0$

13 If S is a closed surface, show that $\iint_S \mathbf{N} \cdot \mathbf{R}\, dS = 3V$, where V is the volume enclosed by S

14 If S is an arbitrary closed surface and $\iint_S \mathbf{N} \cdot \mathbf{F}\, dS = 0$, can we conclude that $\mathbf{F} \equiv 0$? Can we if S is an arbitrary open surface?

15 By applying the divergence theorem to the vector $\phi\mathbf{A}$, where $\mathbf{A}$ is an arbitrary constant vector, show that $\iint_S \phi\mathbf{N}\, dS = \iiint_V \nabla\phi\, dV$ What is $\iint_S \mathbf{N}\, dS$?

16 By applying Stokes' theorem to the vector $\phi\mathbf{A}$, where $\mathbf{A}$ is an arbitrary constant vector, show that $\int_C \phi\, d\mathbf{R} = \iint_S \mathbf{N} \times \nabla\phi\, dS$

17 If S is an open surface, what is $\iint_S \mathbf{N} \times \mathbf{R}\, dS$? (Hint Use the result of Exercise 16)

18 By applying Stokes' theorem to the vector $\mathbf{F} \times \mathbf{A}$, where $\mathbf{A}$ is an arbitrary constant vector, show that $\int_C d\mathbf{R} \times \mathbf{F} = \iint_S (\mathbf{N} \times \nabla) \times \mathbf{F}\, dS$ What is $\int_C d\mathbf{R} \times \mathbf{R}$?

19 Verify the divergence theorem for the function $2xz\mathbf{i} + yz\mathbf{j} + z^2\mathbf{k}$ over the upper half of the sphere $x^2 + y^2 + z^2 = a^2$.

20 Verify the divergence theorem for the function $y\mathbf{i} + x\mathbf{j} + z^2\mathbf{k}$ over the cylindrical region bounded by $x^2 + y^2 = a^2$, $z = 0$, and $z = a$

21 Verify the divergence theorem for the function $x^2\mathbf{i} + z\mathbf{j} + yz\mathbf{k}$ over the cube whose vertices are (0,0,0), (1,0,0), (1,1,0), (0,1,0), (0,0,1), (1,0,1), (1,1,1), and (0,1,1)

22 Verify Stokes' theorem for the function $xy\mathbf{i} + yz\mathbf{j} + z^2\mathbf{k}$ over the cube described in Exercise 21 if the face of the cube in the xy-plane is missing

23 What is the surface integral of the normal component of the curl of the vector $(x + y)\mathbf{i} + (y - x)\mathbf{j} + z^2\mathbf{k}$ over the upper half of the sphere $x^2 + y^2 + z^2 = 1$?

24 If at each point of a surface S the vector $\mathbf{F}(x,y,z)$ is perpendicular to S, prove that the curl of $\mathbf{F}$ either vanishes identically or is everywhere tangent to S (Hint Apply Stokes' theorem to $\mathbf{F}$ over the portion of S bounded by an arbitrary closed curve on S)

25 If at each point of a closed surface S the vector $\mathbf{F}(x,y,z)$ is perpendicular to S, prove that $\iiint_V \nabla \times \mathbf{F}\, dV = 0$ (Hint· Use the result of Example 1)

26 If **A** is an arbitrary constant vector, show that $\iint_S \mathbf{N} \times (\mathbf{A} \times \mathbf{R})\, dS = 2V\mathbf{A}$, where V is the volume bounded by the closed surface S (Hint Use the result of Example 1)

27 Show that $\iiint_V \left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2}\right) dV = \iint_S \frac{d\phi}{dn}\, dS$, where $\frac{d\phi}{dn}$ is the directional derivative of ϕ in the direction of the outer normal to the closed surface S which bounds the volume V

28 If $\phi(x,y,z)$ is a solution of Laplace's equation, show that

$$\iiint_V \left[\left(\frac{\partial \phi}{\partial x}\right)^2 + \left(\frac{\partial \phi}{\partial y}\right)^2 + \left(\frac{\partial \phi}{\partial z}\right)^2\right] dV = \iint_S \phi \frac{d\phi}{dn}\, dS$$

where $\frac{d\phi}{dn}$ is the directional derivative of ϕ in the direction of the outer normal to the closed surface S Hence show also that $\iint_S \phi \frac{d\phi}{dn}\, dS > 0$ if ϕ is a solution of Laplace's equation

29 Extend Gauss' theorem to the case in which O lies *on* the surface S

30 Show that although the function

$$\mathbf{F} = \frac{-y}{x^2 + y^2}\mathbf{i} + \frac{x}{x^2 + y^2}\mathbf{j} + \mathbf{k}$$

is continuous and equal to the gradient of

$$\phi(x,y,z) = \tan^{-1}\frac{y}{x} + z$$

at all points of the region between the two cylinders

$$x^2 + y^2 = \tfrac{1}{4} \qquad \text{and} \qquad x^2 + y^2 = 4$$

the integral $\int \mathbf{F} \cdot d\mathbf{R}$ is not independent of the path in this region [Hint Take A to be $(-1,0,0)$ and B to be $(1,0,0)$, and compute $\int_A^B \mathbf{F} \cdot d\mathbf{R}$ along the upper and lower arcs of the circle $x^2 + y^2 = 1$, $z = 0$]

12.6

Further applications

One of the most important uses of vector analysis is in the concise formulation of physical laws and the derivation of other results from those laws As a first example of this sort, we shall develop the concept of *potential* and obtain the partial differential equation satisfied by the gravitational potential

To do this, let us suppose that we have a **field of force** of some kind, or, in other words, let us consider a region of space in which at every point a force vector **F** is defined The field might, for instance, be **gravitational,** in which case $\mathbf{F}(x,y,z)$ would be the force acting on a unit mass at the general point P (x,y,z) because of the attraction of other masses present in the region On the other hand, the field might be **electrostatic,** in which case $\mathbf{F}(x,y,z)$ would be the force acting on a unit charge at the general point $P.(x,y,z)$ because of the attraction or repulsion of other charges present in the region Or the field might be **magnetic,** in which case $\mathbf{F}(x,y,z)$ would be the force acting on a unit magnetic pole situated at the point P (x,y,z). In any case, the force $\mathbf{F}$ experienced by a unit test body of the appropriate nature is called the **field intensity.**

Now, the amount of work that must be done when a unit test body is moved along an arbitrary curve in the force field defined by a vector function $\mathbf{F}$ is the line integral of the tangential component of $\mathbf{F}$; that is,

$$W = \int \mathbf{F} \cdot d\mathbf{R}$$

If there is no dissipation of energy through friction or similar effects, then, according to the law of the conservation of energy, this integral must be zero around every closed path, and, hence, by Theorem 6, Sec 12 5, it must be independent of the path between any given points A and B Fields for which this is the case are said to be **conservative.** Furthermore, according to Theorem 6, Sec 12 5, it is clear that in a conservative field the force vector $\mathbf{F}$ is the gradient of the scalar function

$$\phi(x,y,z) = \int_{P_0}^{P} \mathbf{F} \cdot d\mathbf{R}$$

The function ϕ is called the **potential function*** of the field In most problems, the masses or charges which produce $\mathbf{F}$ are given, and it is required to find $\mathbf{F}$ itself Since $\mathbf{F} = \nabla\phi$, it is clear that knowing ϕ is equivalent to knowing $\mathbf{F}$, and, hence, the determination of ϕ is of prime importance in most field problems

Assuming, for definiteness, that we are dealing with a gravitational field, let $\mathbf{F}$ be the field intensity at a general point P (x,y,z), and let $\Delta\mathbf{F}$ be the contribution to $\mathbf{F}$ due to the infinitesimal mass Δm_1 in an infinitesimal volume $\Delta V_1 = \Delta x_1 \Delta y_1 \Delta z_1$ enclosing the point P_1 (x_1,y_1,z_1) According to Newton's law of universal gravitation, $\Delta\mathbf{F}$ is a vector whose magnitude is

$$\Delta F = k \frac{1 \quad \Delta m_1}{r^2}$$

where $\qquad r^2 = (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2$

and whose direction is opposite to that of the vector

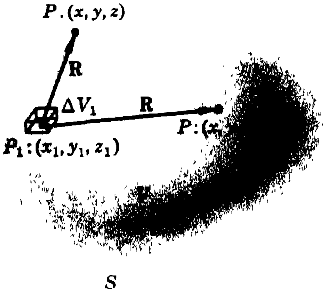
$$\mathbf{R} = (x - x_1)\mathbf{i} + (y - y_1)\mathbf{j} + (z - z_1)\mathbf{k}$$

extending from P_1 to P (Fig 12 30) In other words, if units are so chosen that the constant in Newton's law is equal to unity, the

---

* Many writers define the potential to be $\int_P^{P_0} \mathbf{F} \cdot d\mathbf{R}$, in which case $\mathbf{F} = -\nabla\phi$ In particular, P_0 is often taken to be infinitely distant, so that $\phi = \int_P^{\infty} \mathbf{F} \cdot d\mathbf{R}$

FIGURE 12.30 Figure used in calculating the potential at a point P due to the material in a volume element ΔV_1



field intensity at P due to the infinitesimal mass Δm_1 at P_1 is

$$\Delta \mathbf{F} = -\frac{\Delta m_1}{r^2}\,\frac{\mathbf{R}}{r} = -\rho(x_1,y_1,z_1)\,\Delta V_1\,\frac{\mathbf{R}}{r^3} \tag{1}$$

where $\rho(x_1,y_1,z_1)$ is the density of the material at the point P_1

Now let S be an arbitrary closed regular surface bounding a volume V, and let I denote the integral over S of the normal component of the force due to all the attracting material in the field. By definition, since $\mathbf{F} = \nabla\phi$, we have

$$I = \iint_S \mathbf{N}\cdot\mathbf{F}\,dS = \iint_S \mathbf{N}\cdot\nabla\phi\,dS \tag{2}$$

However, I can also be computed by first determining the part ΔI of it due to the material within ΔV_1 and then taking all the material in the field into account by integration From this point of view we have, from (1) and (2),

$$\begin{aligned}\Delta I = \iint_S \mathbf{N}\cdot\Delta\mathbf{F}\,dS &= -\iint_S [\rho(x_1,y_1,z_1)\,\Delta V_1]\mathbf{N}\cdot\frac{\mathbf{R}}{r^3}\,dS \\ &= -\rho(x_1,y_1,z_1)\,\Delta V_1 \iint_S \mathbf{N}\cdot\frac{\mathbf{R}}{r^3}\,dS\end{aligned}$$

since x_1, y_1, z_1 are constant with respect to the x,y,z-integration over S The last integral can, of course, be evaluated by Gauss' theorem (Theorem 3, Sec 12.5) Specifically, if the origin of $\mathbf{R}$, namely, the point P_1 (x_1,y_1,z_1), is within S, the value of the integral is 4π, otherwise the value of the integral is 0 Hence,

$$\Delta I = \begin{cases} -4\pi\rho(x_1,y_1,z_1)\,\Delta V_1 & \Delta V_1 \text{ within } S \\ 0 & \Delta V_1 \text{ outside } S \end{cases}$$

and, therefore, in computing I it is necessary to integrate only over the volume V bounded by S. Doing this, we find

$$I = \int dI = -4\pi \iiint_V \rho(x_1,y_1,z_1)\,dV_1$$

or, since x_1, y_1, z_1 are just dummy variables,

$$I = -4\pi \iiint_V \rho(x,y,z)\,dV \tag{3}$$

Equating the two expressions (2) and (3) which we now have for I, we get

$$\iint_S \mathbf{N}\cdot\nabla\phi\,dS = -4\pi \iiint_V \rho(x,y,z)\,dV$$

If we now apply the divergence theorem to the integral on the left, we have

$$\iiint_V \nabla \cdot (\nabla\phi)\, dV = -4\pi \iiint_V \rho(x,y,z)\, dV$$

or $$\iiint_V [\nabla^2\phi + 4\pi\rho(x,y,z)]\, dV = 0$$

Since this holds for any arbitrary volume V, it follows that the integrand must vanish identically,* and, therefore, that

(4) $$\nabla^2\phi = -4\pi\rho(x,y,z)$$

This is **Poisson's equation,**† and we have thus shown that *in regions occupied by matter, the gravitational potential satisfies Poisson's equation* In empty space $\rho(x,y,z) = 0$, and thus *in empty space the gravitational potential satisfies Laplace's equation*

(5) $$\nabla^2\phi = 0$$

Results similar to these hold for the electrostatic and magnetic potentials.

As a second example of the use of vector analysis in formulating physical laws in mathematical terms, we shall now derive *Maxwell's equations*‡ for electric and magnetic fields To do this we shall have to work with the vector quantities·

$\mathbf{E}$ = electric intensity
$\mathbf{H}$ = magnetic intensity
$\mathbf{D} = \epsilon\mathbf{E}$ = electric flux density
$\mathbf{B} = \mu\mathbf{H}$ = magnetic flux density
$\mathbf{J}$ = current density

and the scalars

ϵ = permittivity
μ = permeability
σ = conductivity
Q = charge density
$q = \iiint_V Q\, dV$ = total charge within V
$\phi = \iint_S \mathbf{N} \cdot \mathbf{B}\, dS$ = total magnetic flux passing through S
$i = \iint_S \mathbf{N} \cdot \mathbf{J}\, dS$ = total current flowing through S

---

* Suppose that this is not the case, and let P_0 be a point at which the integrand does not vanish Then, if $\rho(x,y,z)$ and $\nabla^2\phi$ are continuous (as we have implicitly assumed), it follows that, throughout some sufficiently small three-dimensional region V_0 enclosing P_0, the integrand has everywhere the same sign it has at P_0 Integrating over V_0, we then obtain an integral which is not equal to zero, contrary to the fact that the integral has been shown to be zero for every volume V

† Named for the French mathematical physicist Simeon Denis Poisson (1781–1840)

‡ Named for the English mathematical physicist James Clerk Maxwell (1831–1879).

These quantities are connected by a number of equations expressing relations discovered experimentally in the early years of the nineteenth century, chiefly by Michael Faraday (1791–1867). In particular we have **Faraday's law,**

$$\int_C \mathbf{E} \cdot d\mathbf{R} = -\frac{\partial \phi}{\partial t} \tag{6}$$

which asserts that the integral of the tangential component of the electric intensity vector around any closed curve C is equal but opposite in sign to the rate of change of the magnetic flux passing through any surface spanning C, **Ampère's law,**

$$\int_C \mathbf{H} \cdot d\mathbf{R} = i \tag{7}$$

which asserts that the integral of the tangential component of the magnetic intensity vector around any closed curve is equal to the current flowing through any surface spanning C, **Gauss' law for electric fields,**

$$\iint_S \mathbf{N} \cdot \mathbf{D}\, dS = q \tag{8}$$

which asserts that the integral of the normal component of the electric flux density over any closed surface S is equal to the total electric charge enclosed by S; and **Gauss' law for magnetic fields,**

$$\iint_S \mathbf{N} \cdot \mathbf{B}\, dS = 0 \tag{9}$$

which asserts that the total magnetic flux ϕ passing through a closed surface is zero.

If we now apply Stokes' theorem to Faraday's law (6), we have

$$\iint_S \mathbf{N} \cdot \nabla \times \mathbf{E}\, dS = -\frac{\partial \phi}{\partial t}$$

and, substituting for ϕ from its definition in terms of $\mathbf{B}$,

$$\iint_S \mathbf{N} \cdot \nabla \times \mathbf{E}\, dS = -\frac{\partial}{\partial t}\left(\iint_S \mathbf{N} \cdot \mathbf{B}\, dS\right) = -\iint_S \mathbf{N} \cdot \frac{\partial \mathbf{B}}{\partial t}\, dS$$

Since S is an arbitrary surface spanning the arbitrary closed curve C, the last equation can hold only if

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{10}$$

Similarly, by applying Stokes' theorem to Ampère's law (7), we obtain

$$\iint_S \mathbf{N} \cdot \nabla \times \mathbf{H}\, dS = i = \iint_S \mathbf{N} \cdot \mathbf{J}\, dS$$

and again, since S is an arbitrary open surface, we conclude that the vectors being integrated over S must be identical

$$\nabla \times \mathbf{H} = \mathbf{J} \tag{11}$$

Now, as Maxwell was the first to realize, the current density $\mathbf{J}$ consists of two parts, namely, a conduction current density

$$\mathbf{J}_c = \sigma\mathbf{E}$$

due to the flow of electric charges, and a displacement current density

$$\mathbf{J}_d = \frac{\partial \mathbf{D}}{\partial t} = \epsilon\frac{\partial \mathbf{E}}{\partial t}$$

due to the time variation of the electric field Thus,

$$\mathbf{J} = \sigma\mathbf{E} + \epsilon\frac{\partial \mathbf{E}}{\partial t}$$

and (11) becomes

(12) $$\nabla \times \mathbf{H} = \sigma\mathbf{E} + \epsilon\frac{\partial \mathbf{E}}{\partial t}$$

Next we apply the divergence theorem to the first of Gauss' laws (8), getting

$$\iiint_V \nabla \cdot \mathbf{D}\, dV = q = \iiint_V Q\, dV$$

whence, since V is arbitrary,

(13) $$\nabla \cdot \mathbf{D} = Q$$

In the same way, by applying the divergence theorem to Gauss' second law (9), we find that

$$\iiint_V \nabla \cdot \mathbf{B}\, dV = 0$$

and, therefore, since V is arbitrary,

(14) $$\nabla \cdot \mathbf{B} = 0$$

Now, if we take the curl of Eq (10), we obtain

$$\nabla \times (\nabla \times \mathbf{E}) = -\nabla \times \frac{\partial \mathbf{B}}{\partial t} = -\frac{\partial}{\partial t}(\nabla \times \mathbf{B}) = -\mu\frac{\partial}{\partial t}(\nabla \times \mathbf{H})$$

If we expand the term $\nabla \times (\nabla \times \mathbf{E})$ by means of Eq (20), Sec 12 3, the last equation becomes

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E} = -\mu\frac{\partial}{\partial t}(\nabla \times \mathbf{H})$$

and, substituting for $\nabla \times \mathbf{H}$ from (12),

(15) $$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E} = -\mu\frac{\partial}{\partial t}\left(\sigma\mathbf{E} + \epsilon\frac{\partial \mathbf{E}}{\partial t}\right)$$

Now, if the space charge density Q is zero, as it is to a high degree of approximation in both good dielectrics and good conductors, then from (13) and the relation $\mathbf{D} = \epsilon\mathbf{E}$ we see that

$$\nabla \cdot \mathbf{E} = 0$$

Therefore, Eq. (15) reduces to

$$\nabla^2\mathbf{E} = \mu\epsilon\frac{\partial^2\mathbf{E}}{\partial t^2} + \mu\sigma\frac{\partial\mathbf{E}}{\partial t}$$

which is **Maxwell's equation for the electric intensity vector E.**

Similarly, if we take the curl of Eq. (12) we obtain

$$\nabla\times(\nabla\times\mathbf{H}) = \nabla\times\left(\sigma\mathbf{E} + \epsilon\frac{\partial\mathbf{E}}{\partial t}\right)$$

and, expanding the left-hand side,

$$\begin{aligned}\nabla(\nabla\cdot\mathbf{H}) - \nabla^2\mathbf{H} &= \sigma\nabla\times\mathbf{E} + \epsilon\nabla\times\frac{\partial\mathbf{E}}{\partial t}\\ &= \sigma\nabla\times\mathbf{E} + \epsilon\frac{\partial}{\partial t}(\nabla\times\mathbf{E})\end{aligned}$$

Now, substituting for $\nabla\times\mathbf{E}$ from (10), we have

$$\nabla(\nabla\cdot\mathbf{H}) - \nabla^2\mathbf{H} = \sigma\left(-\frac{\partial\mathbf{B}}{\partial t}\right) + \epsilon\left(-\frac{\partial^2\mathbf{B}}{\partial t^2}\right)$$

But $\mathbf{B} = \mu\mathbf{H}$, by definition Hence, (14) implies that $\nabla\cdot\mathbf{H} = 0$; therefore, the last equation reduces to

$$\nabla^2\mathbf{H} = \mu\epsilon\frac{\partial^2\mathbf{H}}{\partial t^2} + \mu\sigma\frac{\partial\mathbf{H}}{\partial t}$$

which is **Maxwell's equation for the magnetic intensity vector H.**

For a perfect dielectric, $\sigma = 0$ Hence, in this case Maxwell's equations reduce to the three-dimensional wave equations

$$\nabla^2\mathbf{E} = \mu\epsilon\frac{\partial^2\mathbf{E}}{\partial t^2} \quad\text{and}\quad \nabla^2\mathbf{H} = \mu\epsilon\frac{\partial^2\mathbf{H}}{\partial t^2}$$

On the other hand, in a good conductor the terms arising from the displacement current, i e., the terms containing the second time derivatives, are negligible, and Maxwell's equations reduce to

$$\nabla^2\mathbf{E} = \mu\sigma\frac{\partial\mathbf{E}}{\partial t} \quad\text{and}\quad \nabla^2\mathbf{H} = \mu\sigma\frac{\partial\mathbf{H}}{\partial t}$$

which are examples of the three-dimensional heat equation

As a final application of the methods of vector analysis, we shall investigate the question of whether or not a solution of the heat equation satisfying prescribed boundary and initial conditions over a given region is necessarily unique In our discussion of boundary value problems in Chap. 8 we proceeded on the assumption that this was the case Nevertheless, examples have been given* of solutions of the one-dimensional heat equation

$$a^2\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

---

* See, for instance, P C Rosenbloom and D V Widder, "A Temperature Function which Vanishes Identically," *Am Math Monthly*, vol 65, p. 607, October, 1958

which possess derivatives of all orders for all values of x and t, satisfy identical initial conditions everywhere on the entire x-axis, and yet are different! Confronted with such a clear-cut failure of intuition, we must regard the uniqueness question as of more than academic interest and any positive result as having important practical significance.

Let us suppose, then, that we are to solve the three-dimensional heat equation

$$a^2 \frac{\partial u}{\partial t} = \nabla^2 u$$

throughout a region V bounded by the closed surface S, subject to the boundary condition

$$u = f(x,y,z,t) \qquad \text{on } S$$

and the initial condition

$$u(x,y,z,0) = g(x,y,z) \qquad \text{throughout } V$$

Furthermore, let us suppose that we have two solutions of the problem, u_1 and u_2, each of which, with its derivatives through the second, is continuous in V

If we define a new function

$$w(x,y,z,t) = u_2(x,y,z,t) - u_1(x,y,z,t)$$

it is clear from the linearity of the heat equation that w also satisfies this equation Moreover, w obviously assumes boundary and initial conditions which are identically zero Finally, w is continuous and differentiable, since it is the difference of two functions with these properties

Now consider the volume integral

(16) $$J(t) = \tfrac{1}{2} \iiint_V w^2(x,y,z,t)\, dV \qquad t \geqq 0$$

Clearly, $J(t)$ is a continuous function which is always equal to or greater than zero, since its integrand is everywhere nonnegative Also, since $w = 0$ when $t = 0$, it follows that $J(0) = 0$ Now

$$J'(t) = \frac{1}{2} \iiint_V 2w \frac{\partial w}{\partial t}\, dV$$

and, thus, since w satisfies the heat equation, we have

(17) $$J'(t) = \frac{1}{a^2} \iiint_V w\nabla^2 w\, dV$$

To this, let us apply Eq (7), Sec. 12.5, with both u and v in the formula taken to be the function w of the present problem Then

(18) $$\iiint_V (w\nabla^2 w + \nabla w \cdot \nabla w)\, dV = \iint_S \mathbf{N} \cdot w\, \nabla w\, dS$$

Since the function w vanishes identically on S, the integral on the right side of (18) is zero, and we have

$$\iiint_V w\nabla^2 w\, dV = - \iiint_V \nabla w \cdot \nabla w\, dV$$

Hence, substituting into (17),

$$J'(t) = -\frac{1}{a^2}\iiint_V \nabla w \cdot \nabla w \, dV$$
$$= -\frac{1}{a^2}\iiint_V \left[\left(\frac{\partial w}{\partial x}\right)^2 + \left(\frac{\partial w}{\partial y}\right)^2 + \left(\frac{\partial w}{\partial z}\right)^2\right] dV$$

which shows that

$$J'(t) \leqq 0 \qquad \text{for } t \geqq 0$$

Now, by the law of the mean,

$$\frac{J(t) - J(0)}{t} = J'(t_1) \qquad 0 < t_1 < t$$

or $\quad J(t) = J(0) + tJ'(t_1) \qquad 0 < t_1 < t$

But we have already verified that $J(0) = 0$. Hence, the last equation reduces to

$$J(t) = tJ'(t_1)$$

which shows that

(19) $$J(t) \leqq 0 \qquad \text{for } t \geqq 0$$

since we have just proved that $J'(t)$ is nonpositive for $t \geqq 0$ However, as we observed earlier, the definition of $J(t)$ shows that

(20) $$J(t) \geqq 0 \qquad \text{for } t \geqq 0$$

The only way in which the inequalities (19) and (20) can simultaneously be fulfilled is for $J(t)$ to be identically zero. But this is possible if and only if the integrand of $J(t)$ vanishes identically Hence,

$$w(x,y,z,t) \equiv u_2(x,y,z,t) - u_1(x,y,z,t) = 0$$

or $\quad u_2(x,y,z,t) = u_1(x,y,z,t)$

Thus *in bounded regions, twice differentiable solutions of the heat equation satisfying prescribed surface and initial temperature conditions are unique*

EXERCISES

1 What is the potential function for a central force field in which the attraction on a particle varies directly as the square of the distance from the origin? inversely as the distance from the origin?

2 What is the potential function of the force field due to uniform rotation about the z-axis?

3 What is the potential function for the gravitational field of a uniform circular disk at any point on the axis of the disk?

4 What is the potential function for the gravitational field of a uniform sphere of radius a and mass M? Show that the attraction of the sphere at a point P a distance r from the center of the sphere is

$$\mathbf{F} = \begin{cases} -\dfrac{M\mathbf{R}}{a^3} & r \leqq a \\ -\dfrac{M\mathbf{R}}{r^3} & r \geqq a \end{cases}$$

5 Show that the electrostatic field intensity at a point P due to a set of charges q_i is equal to

$$\mathbf{E} = -\sum_{i=1}^{n} \frac{q_i}{r_i^3}\mathbf{R}_i$$

where $\mathbf{R}_i$ is the vector from the point P to the point P_i where the charge q_i is located Verify that $\nabla \cdot \mathbf{E} = 0$ in this case

6 Show that the work done in bringing a charge of strength q from infinity to a point at a distance of r_0 from a fixed charge q_0 is qq_0/r_0 Using this result, determine the total energy in the electrostatic field defined by the fixed charges $q_1, q_2, \ldots, q_n$ whose mutual distances are r_{ij}

7 If a **conductor** is defined to be a body in whose interior the electric field is everywhere zero, show that any charge on a conductor must be located entirely on its surface.

8 Let V_1 and V_2 be two regions with respective dielectric constants ϵ_1 and ϵ_2, and let S be the surface of discontinuity which separates them By applying Gauss' law for electric fields to a closed cylindrical surface of infinitesimal height whose bases are parallel to S in the respective media, show that, if there are no charges on S, the normal component of the electric flux density is continuous across S Similarly, by applying Faraday's law to a rectangle of negligible width whose longer sides are parallel to S in the respective media, prove that, if the field is conservative, the tangential component of the electric intensity is continuous across S

9 What is the electric field in the empty space between the perfectly conducting, infinite planes $y = 0$ and $y = l$ if $\mathbf{E}\Big|_{t=0} = \mathbf{i} + \mathbf{k}$ and $\dfrac{\partial \mathbf{E}}{\partial t}\Big|_{t=0} = \mathbf{i} - \mathbf{k}$? (Hint From the nature of the region of the problem and the initial conditions, it is clear that the field has no component in the y-direction and that E_x and E_z are functions only of y)

10 Prove that a solution of the heat equation, possessing continuous second partial derivatives, which takes on prescribed initial values throughout a region V and whose normal derivative takes on prescribed values on the surface S which encloses V is unique

CHAPTER THIRTEEN

Tensor Analysis

13.1

Introduction

In Chap 10 we introduced the concept of a vector as either a $(1,n)$ or an $(n,1)$ matrix, that is, as an ordered set of n quantities. In the last chapter we took a somewhat less abstract point of view and regarded a vector as a quantity which could be represented by a directed line segment Using this interpretation we then developed the algebra and calculus of vectors In doing this, we worked implicitly (and sometimes explicitly) in a rectangular frame of reference, nonetheless, it should be clear that we were dealing with quantities independent of any particular coordinate system For example, though the description of a point, that is, its coordinates, may change from one coordinate system to another, the point is recognizably the same in all coordinate systems. Similarly, although the formula by which it is computed may change, the length of a particular vector must be the same in all coordinate systems.

In this chapter we shall pursue further this idea of invariance and adopt as our fundamental idea of a vector the concept of a quantity invariant under any transformation of coordinates This will lead us to the idea of the covariant and contravariant representation of vectors and, thence, to the highly important concept of a *tensor* Although we cannot undertake a detailed discussion of tensor analysis, we shall undertake to indicate some of its principal features and illustrate the remarkable economy of the tensor notation

13.2

Oblique coordinates

Because of the need to distinguish between what we shall soon refer to as *covariant* and *contravariant vectors*, it is necessary that our notation employ indices not only in the familiar subscript position but in the superscript position as well. In tensor analysis this requirement takes precedence over the usual exponential symbolism, and, henceforth, when we write, say,

$$\xi^\alpha$$

α will be a distinguishing index, like subscripts heretofore, and *not* an exponent If and when we wish to indicate the αth power of a quantity ξ we shall always use parentheses and write

$$(\xi)^\alpha$$

With this convention in mind, and as a relatively simple example of the generalized coordinates we shall subsequently investigate, let us consider a system of coordinates $(\bar{x}^1,\bar{x}^2,\bar{x}^3)$ connected with a system of rectangular coordinates (x^1,x^2,x^3) by the equations

$$\begin{aligned} \bar{x}^1 &= a_{11}x^1 + a_{12}x^2 + a_{13}x^3 \\ \bar{x}^2 &= a_{21}x^1 + a_{22}x^2 + a_{23}x^3 \\ \bar{x}^3 &= a_{31}x^1 + a_{32}x^2 + a_{33}x^3 \end{aligned} \qquad |A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \neq 0$$

or, in matrix form,

(1) $$\bar{X} = AX$$

and

(2) $$X = A^{-1}\bar{X}$$

where, as usual,

$$X = \left\|\begin{matrix} x^1 \\ x^2 \\ x^3 \end{matrix}\right\| \qquad \text{and} \qquad \bar{X}\dagger = \left\|\begin{matrix} \bar{x}^1 \\ \bar{x}^2 \\ \bar{x}^3 \end{matrix}\right\|$$

The locus of points for which $\bar{x}^1 = 0$ is, of course, the plane

$$\pi_1\colon \qquad a_{11}x^1 + a_{12}x^2 + a_{13}x^3 = 0$$

Similarly, the locus of points for which $\bar{x}^2 = 0$ is the plane

$$\pi_2 \qquad a_{21}x^1 + a_{22}x^2 + a_{23}x^3 = 0$$

and the locus of points for which $\bar{x}^3 = 0$ is the plane

$$\pi_3 \qquad a_{31}x^1 + a_{32}x^2 + a_{33}x^3 = 0$$

Clearly, on the line of intersection of π_2 and π_3, both $\bar{x}^2$ and $\bar{x}^3$ are zero and $\bar{x}^1$ alone varies This line can, therefore, be thought of as the $\bar{x}^1$-axis In the same fashion we can identify the line of intersection of π_1 and π_3 as the $\bar{x}^2$-axis, and the line of intersection of π_1 and π_2 as the $\bar{x}^3$-axis (Fig 13.1*a*). Since the point for which $\bar{x}^1 = \bar{x}^2 = \bar{x}^3 = 0$ obviously lies in π_1, π_2, and π_3, it follows that

---

† Not only the new coordinates themselves, but all quantities referred to the new coordinate system we shall consistently denote by overbars Thus, if P is the name of a point described in the original (rectangular) coordinate system by the coordinates (p^1,p^2,p^3), then $\bar{P}$ is the name we shall use for this point thought of as described by the new coordinates $(\bar{p}^1,\bar{p}^2,\bar{p}^3)$ determined by Eq (1)

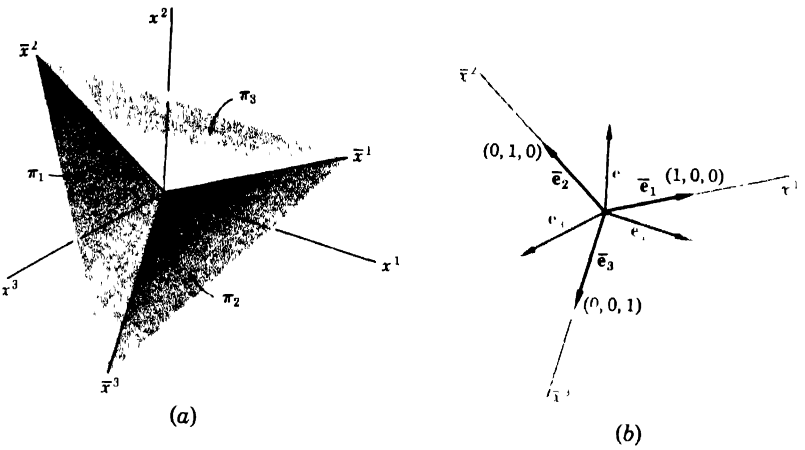


FIGURE 13.1

A rectangular and an oblique coordinate system with their related reference vectors

the $\bar{x}^1$-, $\bar{x}^2$-, $\bar{x}^3$-axes are concurrent. Moreover, since $|A| \neq 0$, these lines are distinct and noncoplanar. In general, however, they will not be mutually perpendicular, and for this reason they are said to be the axes of an **oblique coordinate system.**

Since the $\bar{x}^1$-, $\bar{x}^2$-, and $\bar{x}^3$-axes are noncoplanar, any vector can be expressed as a linear combination of arbitrary reference vectors along the three oblique axes. By analogy with the unit vectors **i**, **j**, **k**, or $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, as we shall now denote them, it might seem natural to choose vectors of unit length for this purpose. However, because the oblique coordinates $\bar{x}^1$, $\bar{x}^2$, $\bar{x}^3$ are *not* distance measures along the oblique axes, as x^1, x^2, x^3 are along the axes of a rectangular coordinate system, it turns out to be more convenient to take the new reference vectors $\bar{\mathbf{e}}_1$, $\bar{\mathbf{e}}_2$, $\bar{\mathbf{e}}_3$ to be, respectively, the vectors from the origin to the points whose oblique coordinates are (1,0,0), (0,1,0), and (0,0,1) (Fig. 13.1*b*).

To determine the lengths of the reference vectors $\bar{\mathbf{e}}_1$, $\bar{\mathbf{e}}_2$, $\bar{\mathbf{e}}_3$ and to obtain the formula for measuring distances in general in oblique coordinates, let us consider the vector $\bar{\mathbf{V}}$† extending from the point $\bar{P}$ whose matrix of oblique coordinates is $\bar{V}_P = \begin{Vmatrix} \bar{p}^1 \\ \bar{p}^2 \\ \bar{p}^3 \end{Vmatrix}$ to the point $\bar{Q}$ whose matrix of oblique coordinates is $\bar{V}_Q = \begin{Vmatrix} \bar{q}^1 \\ \bar{q}^2 \\ \bar{q}^3 \end{Vmatrix}$. From Eq. (2) it follows that the rectangular coordinates of $\bar{P}$ $(= P)$ and

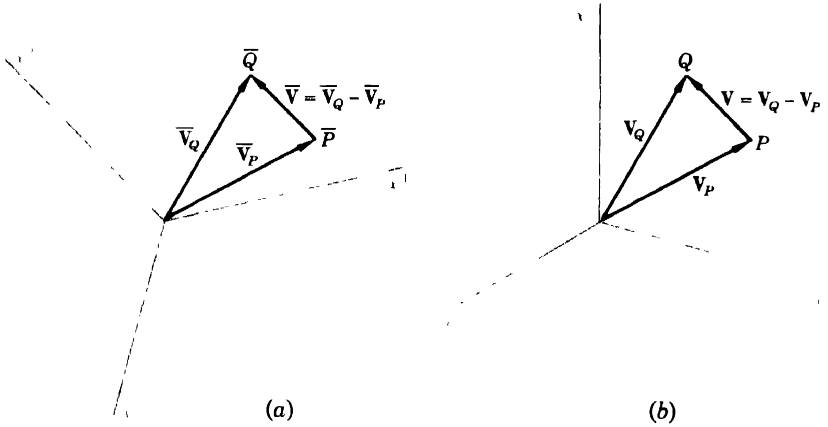
---

† In this chapter we shall use boldface symbols to denote vectors only when we are considering them as directed line segments, as we did in the last chapter. In particular, if $\mathbf{V}$ is a vector considered in the geometric sense, we shall use the symbol V to denote not the length of $\mathbf{V}$ but rather the matrix of the components of $\mathbf{V}$ along the appropriate set of axes.

$\bar{Q}$ ($= Q$) are defined, respectively, by the matrices

$$V_P = A^{-1}\bar{V}_P \qquad \text{and} \qquad V_Q = A^{-1}\bar{V}_Q$$

Hence, in rectangular coordinates, the vector $\bar{\mathbf{V}} = \bar{\mathbf{V}}_Q - \bar{\mathbf{V}}_P$ (Fig. 13.2*a*), defined by the matrix of components $\bar{V} = \bar{V}_Q - \bar{V}_P$,


$\bar{Q}$
$\bar{\mathbf{V}} = \bar{\mathbf{V}}_Q - \bar{\mathbf{V}}_P$
$\bar{\mathbf{V}}_Q$
$\bar{\mathbf{V}}_P$
$\bar{P}$
Q
$\mathbf{V} = \mathbf{V}_Q - \mathbf{V}_P$
$\mathbf{V}_Q$
P
$\mathbf{V}_P$

(*a*) (*b*)

FIGURE 13.2

A vector V represented in each of two coordinate systems

becomes the vector $\mathbf{V} = \mathbf{V}_Q - \mathbf{V}_P$ (Fig. 13.2*b*) defined by the matrix

$$V = V_Q - V_P = A^{-1}\bar{V}_Q - A^{-1}\bar{V}_P = A^{-1}(\bar{V}_Q - \bar{V}_P) = A^{-1}\bar{V}$$

Now, in rectangular coordinates, the square of the length of a vector $\mathbf{V}$ whose matrix of components is V is given by the formula

$$\mathbf{V} \cdot \mathbf{V} = V^T I V = V^T G V$$

where, for later convenience, we have introduced G as another name for the matrix which is I in this case but not in general. Therefore, since we require the length of a given vector to be the same in all coordinate systems, we define the scalar product of a vector with itself in oblique coordinates by the condition that

$$\begin{aligned} \bar{\mathbf{V}} \cdot \bar{\mathbf{V}} = \mathbf{V} \cdot \mathbf{V} = (A^{-1}\bar{V})^T I (A^{-1}\bar{V}) &= \bar{V}^T (A^{-1})^T I A^{-1} \bar{V} \\ &= \bar{V}^T [(A^{-1})^T A^{-1}] \bar{V} \\ &= \bar{V}^T \bar{G} \bar{V} \qquad (3) \end{aligned}$$

Similarly, for distinct vectors $\bar{\mathbf{U}}$ and $\bar{\mathbf{V}}$, we define

$$\begin{aligned} \bar{\mathbf{U}} \cdot \bar{\mathbf{V}} = \mathbf{U} \cdot \mathbf{V} = (A^{-1}\bar{U})^T I (A^{-1}\bar{V}) &= \bar{U}^T (A^{-1})^T I A^{-1} \bar{V} \\ &= \bar{U}^T [(A^{-1})^T A^{-1}] \bar{V} \\ &= \bar{U}^T \bar{G} \bar{V} \qquad (4) \end{aligned}$$

Thus, *the metrical properties of space, which in rectangular coordinates are determined by the identity matrix $I = G$, are in oblique coordinates determined by the matrix $(A^{-1})^T A^{-1} = \bar{G}$, where A is the* *matrix of the transformation* $\bar{X} = AX$ *from rectangular to oblique coordinates.*

Denoting by $\bar{g}_{ij}$ the element in the ith row and jth column of the matrix $\bar{G} = (A^{-1})^T A^{-1}$, it is clear from Eqs (3) and (4) that, for the reference vectors $\bar{\mathbf{e}}_1$, $\bar{\mathbf{e}}_2$, $\bar{\mathbf{e}}_3$ defined by the matrices

$$\bar{e}_1 = \begin{Vmatrix} 1 \\ 0 \\ 0 \end{Vmatrix} \qquad \bar{e}_2 = \begin{Vmatrix} 0 \\ 1 \\ 0 \end{Vmatrix} \qquad \bar{e}_3 = \begin{Vmatrix} 0 \\ 0 \\ 1 \end{Vmatrix} \qquad \text{we have}$$

(5) $$\bar{\mathbf{e}}_i \cdot \bar{\mathbf{e}}_j = \bar{g}_{ij}$$

In particular, the lengths of $\bar{\mathbf{e}}_1$, $\bar{\mathbf{e}}_2$, and $\bar{\mathbf{e}}_3$ are, respectively,

$$|\bar{\mathbf{e}}_1| = \sqrt{\bar{g}_{11}} \qquad |\bar{\mathbf{e}}_2| = \sqrt{\bar{g}_{22}} \qquad |\bar{\mathbf{e}}_3| = \sqrt{\bar{g}_{33}}$$

In other words, *the length of* $\bar{\mathbf{e}}_i$ *is such that, if* $\bar{\mathbf{R}}$ *is the vector extending along the* $\bar{x}^i$*-axis from the origin to the point for which* $\bar{x}^i = \bar{a}^i$, *then the relation* $|\bar{\mathbf{R}}| = |\bar{a}^i|\ |\bar{\mathbf{e}}_i|$ *holds*

In a rectangular coordinate system a unique set of directions for the reference vectors is clearly identified by the axes of the system. In oblique coordinates this is not the case; for, although the oblique axes certainly define a set of directions in which reference vectors can naturally be chosen, there is another set distinct from the first which is also intrinsic in the system, namely, the directions perpendicular to the coordinate planes π_1, π_2, π_3 As base vectors in these directions it is customary to take vectors $\bar{\mathbf{e}}^1$, $\bar{\mathbf{e}}^2$, $\bar{\mathbf{e}}^3$ defined by the conditions

(6) $$\bar{\mathbf{e}}^i \cdot \bar{\mathbf{e}}_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

For $i \neq j$ these relations fix the directions of the new reference vectors, and for $i = j$ they determine their lengths and sense. The vectors $\bar{\mathbf{e}}^1$, $\bar{\mathbf{e}}^2$, $\bar{\mathbf{e}}^3$ are said to form a set **reciprocal** to the set $\bar{\mathbf{e}}_1$, $\bar{\mathbf{e}}_2$, $\bar{\mathbf{e}}_3$, and vice versa* (Fig 13 3) From their definition it is

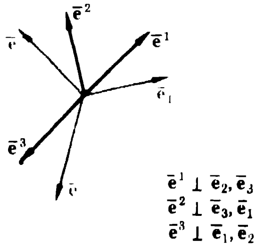

$\bar{\mathbf{e}}^2$
$\bar{\mathbf{e}}$
$\bar{\mathbf{e}}^1$
$\bar{\mathbf{e}}_1$
$\bar{\mathbf{e}}^3$
$\bar{\mathbf{e}}$
$\bar{\mathbf{e}}^1 \perp \bar{\mathbf{e}}_2, \bar{\mathbf{e}}_3$
$\bar{\mathbf{e}}^2 \perp \bar{\mathbf{e}}_3, \bar{\mathbf{e}}_1$
$\bar{\mathbf{e}}^3 \perp \bar{\mathbf{e}}_1, \bar{\mathbf{e}}_2$

FIGURE 13 3 The base vectors $\bar{\mathbf{e}}_1$, $\bar{\mathbf{e}}_2$, $\bar{\mathbf{e}}_3$ and the reciprocal base vectors $\bar{\mathbf{e}}^1$, $\bar{\mathbf{e}}^2$, $\bar{\mathbf{e}}^3$ in an oblique coordinate system.

---

* It is evident that in rectangular coordinates the set of base vectors and the set of reciprocal vectors are the same, that is, $\mathbf{i} \equiv \mathbf{e}_1 = \mathbf{e}^1$, $\mathbf{j} \equiv \mathbf{e}_2 = \mathbf{e}^2$, $\mathbf{k} \equiv \mathbf{e}_3 = \mathbf{e}^3$ It is for this reason that the concept of reciprocal sets of vectors was not introduced in the last chapter (except in Exercise 23, Sec. 12.1).

clear that $\hat{\mathbf{e}}^1, \hat{\mathbf{e}}^2, \hat{\mathbf{e}}^3$ are noncoplanar and, hence, can be used as a basis for the representation of any vector. Thus, when we use oblique coordinates, any vector $\bar{\mathbf{V}}$ has two different but equally natural representations. It can be expressed as a linear combination of the base vectors $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3$, or it can be expressed as a linear combination of the vectors of the reciprocal set $\hat{\mathbf{e}}^1, \hat{\mathbf{e}}^2, \hat{\mathbf{e}}^3$

In particular, the vectors in each of the sets $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3$ and $\hat{\mathbf{e}}^1, \hat{\mathbf{e}}^2, \hat{\mathbf{e}}^3$ must be expressible as linear combinations of the vectors in the other set Specifically, if we write

$$\begin{aligned}
\hat{\mathbf{e}}_1 &= \mu_{11}\hat{\mathbf{e}}^1 + \mu_{12}\hat{\mathbf{e}}^2 + \mu_{13}\hat{\mathbf{e}}^3\\
\hat{\mathbf{e}}_2 &= \mu_{21}\hat{\mathbf{e}}^1 + \mu_{22}\hat{\mathbf{e}}^2 + \mu_{23}\hat{\mathbf{e}}^3\\
\hat{\mathbf{e}}_3 &= \mu_{31}\hat{\mathbf{e}}^1 + \mu_{32}\hat{\mathbf{e}}^2 + \mu_{33}\hat{\mathbf{e}}^3
\end{aligned}$$

and then form the scalar product of each side of the ith equation with $\hat{\mathbf{e}}_j$, we obtain

$$\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \mu_{i1}\hat{\mathbf{e}}^1 \cdot \hat{\mathbf{e}}_j + \mu_{i2}\hat{\mathbf{e}}^2 \cdot \hat{\mathbf{e}}_j + \mu_{i3}\hat{\mathbf{e}}^3 \cdot \hat{\mathbf{e}}_j$$

Hence, using Eqs (5) and (6), we find

$$\bar{g}_{ij} = \mu_{ij}$$

and, therefore,

$$\begin{aligned}
\hat{\mathbf{e}}_1 &= \bar{g}_{11}\hat{\mathbf{e}}^1 + \bar{g}_{12}\hat{\mathbf{e}}^2 + \bar{g}_{13}\hat{\mathbf{e}}^3\\
\hat{\mathbf{e}}_2 &= \bar{g}_{21}\hat{\mathbf{e}}^1 + \bar{g}_{22}\hat{\mathbf{e}}^2 + \bar{g}_{23}\hat{\mathbf{e}}^3\\
\hat{\mathbf{e}}_3 &= \bar{g}_{31}\hat{\mathbf{e}}^1 + \bar{g}_{32}\hat{\mathbf{e}}^2 + \bar{g}_{33}\hat{\mathbf{e}}^3
\end{aligned}\tag{7}$$

If we define the matrices

$$\bar{V}_e = \begin{Vmatrix} \hat{\mathbf{e}}_1 \\ \hat{\mathbf{e}}_2 \\ \hat{\mathbf{e}}_3 \end{Vmatrix} \qquad \text{and} \qquad \bar{V}^e = \begin{Vmatrix} \hat{\mathbf{e}}^1 \\ \hat{\mathbf{e}}^2 \\ \hat{\mathbf{e}}^3 \end{Vmatrix}$$

Eq (7) can be written more compactly in the form

$$\bar{V}_e = \bar{G}\bar{V}^e \tag{8}$$

from which it follows that

$$\bar{V}^e = \bar{G}^{-1}\bar{V}_e \tag{9}$$

or

$$\begin{aligned}
\hat{\mathbf{e}}^1 &= \bar{g}^{11}\hat{\mathbf{e}}_1 + \bar{g}^{12}\hat{\mathbf{e}}_2 + \bar{g}^{13}\hat{\mathbf{e}}_3\\
\hat{\mathbf{e}}^2 &= \bar{g}^{21}\hat{\mathbf{e}}_1 + \bar{g}^{22}\hat{\mathbf{e}}_2 + \bar{g}^{23}\hat{\mathbf{e}}_3\\
\hat{\mathbf{e}}^3 &= \bar{g}^{31}\hat{\mathbf{e}}_1 + \bar{g}^{32}\hat{\mathbf{e}}_2 + \bar{g}^{33}\hat{\mathbf{e}}_3
\end{aligned}\tag{10}$$

where $\bar{g}^{ij}$ is the element in the ith row and jth column of $\bar{G}^{-1}$; that is,

$$g^{ij} = \frac{\bar{G}_{ji}}{|\bar{G}|}$$

Of course, since $\bar{G} = \|\bar{g}_{ij}\| = (A^{-1})^T A^{-1}$ is symmetric, so is its inverse $\bar{G}^{-1} = \|\bar{g}^{ij}\| = AA^T$ From (10) and (6) it follows immediately that

(11) $$\bar{\mathbf{e}}^i \cdot \bar{\mathbf{e}}^j = \bar{g}^{ij}$$

Thus, *in oblique coordinates the metrical properties of space, which are determined by the matrix* $\bar{G} = \|\bar{g}_{ij}\| = (A^{-1})^T A^{-1}$ *if vectors are represented in terms of the base vectors* $\bar{\mathbf{e}}_1$, $\bar{\mathbf{e}}_2$, $\bar{\mathbf{e}}_3$ *are determined equally well by the inverse matrix* $\bar{G}^{-1} = \|\bar{g}^{ij}\| = AA^T$ *if vectors are represented in terms of the reciprocal base vectors* $\bar{\mathbf{e}}^1$, $\bar{\mathbf{e}}^2$, $\bar{\mathbf{e}}^3$.

It is also instructive to consider the representation of the vectors $\mathbf{i} = \mathbf{e}_1 = \mathbf{e}^1$, $\mathbf{j} = \mathbf{e}_2 = \mathbf{e}^2$, $\mathbf{k} = \mathbf{e}_3 = \mathbf{e}^3$ in terms of the vectors $\bar{\mathbf{e}}_1$, $\bar{\mathbf{e}}_2$, $\bar{\mathbf{e}}_3$ and $\bar{\mathbf{e}}^1$, $\bar{\mathbf{e}}^2$, $\bar{\mathbf{e}}^3$, and vice versa. Specifically, since $\bar{\mathbf{e}}_1, \bar{\mathbf{e}}_2, \bar{\mathbf{e}}_3$ are, respectively, the vectors from the origin $\bar{O}$ $(= O)$ to the points whose oblique coordinates are (1,0,0), (0,1,0), and (0,0,1) and since, from the transformation equation $X = A^{-1}\bar{X}$, these points have rectangular coordinates (a^{11},a^{21},a^{31}), (a^{12},a^{22},a^{32}), and (a^{13},a^{23},a^{33}), where $a^{ij} = A_{ji}/|A|$ is the element in the ith row and jth column of the matrix A^{-1}, it follows that

(12)
$$\begin{aligned}\bar{\mathbf{e}}_1 &= a^{11}\mathbf{e}_1 + a^{21}\mathbf{e}_2 + a^{31}\mathbf{e}_3\\ \bar{\mathbf{e}}_2 &= a^{12}\mathbf{e}_1 + a^{22}\mathbf{e}_2 + a^{32}\mathbf{e}_3\\ \bar{\mathbf{e}}_3 &= a^{13}\mathbf{e}_1 + a^{23}\mathbf{e}_2 + a^{33}\mathbf{e}_3\end{aligned}$$

or, introducing the matrices

$$\bar{V}_e = \begin{Vmatrix}\bar{\mathbf{e}}_1\\ \bar{\mathbf{e}}_2\\ \bar{\mathbf{e}}_3\end{Vmatrix} \qquad \text{and} \qquad V_e = \begin{Vmatrix}\mathbf{i}\\ \mathbf{j}\\ \mathbf{k}\end{Vmatrix} = \begin{Vmatrix}\mathbf{e}_1\\ \mathbf{e}_2\\ \mathbf{e}_3\end{Vmatrix} = \begin{Vmatrix}\mathbf{e}^1\\ \mathbf{e}^2\\ \mathbf{e}^3\end{Vmatrix} = V^e$$

(13) $$\bar{V}_e = (A^{-1})^T V_e = (A^T)^{-1} V_e$$

Either in the same fashion or directly from (13), we obtain

(14) $$V_e = A^T \bar{V}_e$$

that is,

(15)
$$\begin{aligned}\mathbf{e}_1 &= a_{11}\bar{\mathbf{e}}_1 + a_{21}\bar{\mathbf{e}}_2 + a_{31}\bar{\mathbf{e}}_3\\ \mathbf{e}_2 &= a_{12}\bar{\mathbf{e}}_1 + a_{22}\bar{\mathbf{e}}_2 + a_{32}\bar{\mathbf{e}}_3\\ \mathbf{e}_3 &= a_{13}\bar{\mathbf{e}}_1 + a_{23}\bar{\mathbf{e}}_2 + a_{33}\bar{\mathbf{e}}_3\end{aligned}$$

as the equations expressing $\mathbf{i} = \mathbf{e}_1$, $\mathbf{j} = \mathbf{e}_2$, $\mathbf{k} = \mathbf{e}_3$ in terms of the base vectors $\bar{\mathbf{e}}_1$, $\bar{\mathbf{e}}_2$, $\bar{\mathbf{e}}_3$ of the oblique system.

To obtain the equations relating $\bar{\mathbf{e}}^1$, $\bar{\mathbf{e}}^2$, $\bar{\mathbf{e}}^3$ and $\mathbf{i} = \mathbf{e}^1$, $\mathbf{j} = \mathbf{e}^2$, $\mathbf{k} = \mathbf{e}^3$, we begin with the relation (9), i e, $\bar{V}^e = \bar{G}^{-1}\bar{V}_e$ From this, using (13) and the fact that $\bar{G}^{-1} = AA^T$ and $V_e = V^e$, we have

(16) $$\bar{V}^e = (AA^T)(A^T)^{-1}V^e = AV^e$$

that is,

$$\begin{aligned} \bar{\mathbf{e}}^1 &= a_{11}\mathbf{e}^1 + a_{12}\mathbf{e}^2 + a_{13}\mathbf{e}^3 \\ \bar{\mathbf{e}}^2 &= a_{21}\mathbf{e}^1 + a_{22}\mathbf{e}^2 + a_{23}\mathbf{e}^3 \\ \bar{\mathbf{e}}^3 &= a_{31}\mathbf{e}^1 + a_{32}\mathbf{e}^2 + a_{33}\mathbf{e}^3 \end{aligned} \tag{17}$$

Solving (16) for V^c, we have, of course,

$$V^c = A^{-1}\bar{V}^c \tag{18}$$

or

$$\begin{aligned} \mathbf{e}^1 &= a^{11}\bar{\mathbf{e}}^1 + a^{12}\bar{\mathbf{e}}^2 + a^{13}\bar{\mathbf{e}}^3 \\ \mathbf{e}^2 &= a^{21}\bar{\mathbf{e}}^1 + a^{22}\bar{\mathbf{e}}^2 + a^{23}\bar{\mathbf{e}}^3 \\ \mathbf{e}^3 &= a^{31}\bar{\mathbf{e}}^1 + a^{32}\bar{\mathbf{e}}^2 + a^{33}\bar{\mathbf{e}}^3 \end{aligned} \tag{19}$$

Suppose now that we have a vector

$$\mathbf{V} = \mathbf{V}^r = u\mathbf{i} + v\mathbf{j} + w\mathbf{k} \equiv v^1\mathbf{e}_1 + v^2\mathbf{e}_2 + v^3\mathbf{e}_3 \equiv v_1\mathbf{e}^1 + v_2\mathbf{e}^2 + v_3\mathbf{e}^3 = \mathbf{V}_r$$

where, since $\mathbf{V}$ is given in a rectangular coordinate system, $\mathbf{e}_i = \mathbf{e}^i$ and $\mathbf{V}^r = \mathbf{V}_r$. If we express $\mathbf{V} \equiv \mathbf{V}^r$ in terms of the base vectors $\bar{\mathbf{e}}_1, \bar{\mathbf{e}}_2, \bar{\mathbf{e}}_3$ of the oblique system by means of (15), we obtain, after collecting terms,

$$\begin{aligned} \bar{\mathbf{V}}^r &= (v^1a_{11} + v^2a_{12} + v^3a_{13})\bar{\mathbf{e}}_1 + (v^1a_{21} + v^2a_{22} + v^3a_{23})\bar{\mathbf{e}}_2 \\ &\qquad\qquad + (v^1a_{31} + v^2a_{32} + v^3a_{33})\bar{\mathbf{e}}_3 \\ &= \bar{v}^1\bar{\mathbf{e}}_1 + \bar{v}^2\bar{\mathbf{e}}_2 + \bar{v}^3\bar{\mathbf{e}}_3 \end{aligned} \tag{20}$$

Similarly, if we express $\mathbf{V} \equiv \mathbf{V}_r$ in terms of the reciprocal base vectors $\bar{\mathbf{e}}^1, \bar{\mathbf{e}}^2, \bar{\mathbf{e}}^3$ by means of (19), we obtain the representation

$$\begin{aligned} \bar{\mathbf{V}}_r &= (v_1a^{11} + v_2a^{21} + v_3a^{31})\bar{\mathbf{e}}^1 + (v_1a^{12} + v_2a^{22} + v_3a^{32})\bar{\mathbf{e}}^2 \\ &\qquad\qquad + (v_1a^{13} + v_2a^{23} + v_3a^{33})\bar{\mathbf{e}}^3 \\ &= \bar{v}_1\bar{\mathbf{e}}^1 + \bar{v}_2\bar{\mathbf{e}}^2 + \bar{v}_3\bar{\mathbf{e}}^3 \end{aligned} \tag{21}$$

Thus, when $\mathbf{V}$ is transformed from its representation in terms of the base vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ to the corresponding representation in terms of the base vectors $\bar{\mathbf{e}}_1, \bar{\mathbf{e}}_2, \bar{\mathbf{e}}_3$, the components of $\mathbf{V} \equiv \mathbf{V}^r$ transform according to the law

$$\bar{v}^i = a_{i1}v^1 + a_{i2}v^2 + a_{i3}v^3 \tag{22}$$

or

$$\bar{V}^r = AV^r \tag{23}$$

Likewise, when $\mathbf{V} \equiv \mathbf{V}_r$ is transformed from its representation in terms of the base vectors $\mathbf{e}^1, \mathbf{e}^2, \mathbf{e}^3$ to its corresponding representation in terms of the reciprocal base vectors $\bar{\mathbf{e}}^1, \bar{\mathbf{e}}^2, \bar{\mathbf{e}}^3$, its components transform according to the law

$$\bar{v}_i = a^{1i}v_1 + a^{2i}v_2 + a^{3i}v_3 \tag{24}$$

or

$$\bar{V}_r = (A^{-1})^T V_r \tag{25}$$

Equations (24) and (25) have exactly the same form as Eqs. (12) and (13) for the transformation of the base vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$; for this reason, the representation of **V** in terms of the reciprocal base vectors is called the **covariant** representation* of **V**. On the other hand, Eqs (22) and (23) have the form of Eqs. (16) and (17) for the transformation of the reciprocal base vectors $\mathbf{e}^1$, $\mathbf{e}^2$, $\mathbf{e}^3$, for this reason, the representation of **V** in terms of the base vectors themselves is called the **contravariant** representation† of **V**.

From Eq (1) it is clear that

$$a_{ij} = \frac{\partial \bar{x}^i}{\partial x^j}$$

and from Eq (2) it is clear that

$$a^{ij} = \frac{\partial x^i}{\partial \bar{x}^j}$$

There is no particular reason for introducing this notation in the study of oblique coordinates, but it may be helpful as a preparation for the work of the next section on generalized coordinates to rewrite some of the important formulas of this section in terms of the partial derivatives of the transformation equations

For the matrix of the transformation and its inverse we have, respectively,

$$A = \|a_{ij}\| = \left\| \frac{\partial \bar{x}^i}{\partial x^j} \right\| \qquad \text{and} \qquad A^{-1} = \|a^{ij}\| = \left\| \frac{\partial x^i}{\partial \bar{x}^j} \right\|$$

For the general element of the matrix $\bar{G} = (A^{-1})^T A^{-1}$, which in oblique coordinates defines the metrical properties of space, we have

$$\bar{g}_{ij} = \sum_k a^{ki}a^{kj} = \sum_k \frac{\partial x^k}{\partial \bar{x}^i}\frac{\partial x^k}{\partial \bar{x}^j}$$

or, inserting the factor g_{kl} which of course is 1 if $k = l$ and 0 if $k \neq l$,

(26)
$$\bar{g}_{ij} = \sum_{k,l} g_{kl}\frac{\partial x^k}{\partial \bar{x}^i}\frac{\partial x^l}{\partial \bar{x}^j}$$

Likewise, for the matrix $\bar{G}^{-1} = \|\bar{g}^{ij}\| = AA^T$, we have

$$\bar{g}^{ij} = \sum_k a_{ik}a_{jk} = \sum_k \frac{\partial \bar{x}^i}{\partial x^k}\frac{\partial \bar{x}^j}{\partial x^k}$$

or, inserting $g^{kl} \equiv g_{kl}$,

(27)
$$\bar{g}^{ij} = \sum_{k,l} g^{kl}\frac{\partial \bar{x}^i}{\partial x^k}\frac{\partial \bar{x}^j}{\partial x^l}$$

---

* Co- = *with* or *alike*

† Contra- = *against* or *opposite to*.

For the relations connecting the base vectors $\bar{\mathbf{e}}_1$, $\bar{\mathbf{e}}_2$, $\bar{\mathbf{e}}_3$ and the vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, we have from (12) and (15),

$$\bar{\mathbf{e}}_i = \sum_k a^{ki}\mathbf{e}_k = \sum_k \mathbf{e}_k \frac{\partial x^k}{\partial \bar{x}^i} \tag{28}$$

and

$$\mathbf{e}_k = \sum_i a_{ik}\bar{\mathbf{e}}_i = \sum_i \bar{\mathbf{e}}_i \frac{\partial \bar{x}^i}{\partial x^k} \tag{29}$$

For the relations connecting the reciprocal base vectors $\bar{\mathbf{e}}^1$, $\bar{\mathbf{e}}^2$, $\bar{\mathbf{e}}^3$ and the vectors $\mathbf{e}^1$, $\mathbf{e}^2$, $\mathbf{e}^3$, we have from (17) and (19),

$$\bar{\mathbf{e}}^i = \sum_k a_{ik}\mathbf{e}^k = \sum_k \mathbf{e}^k \frac{\partial \bar{x}^i}{\partial x^k} \tag{30}$$

and

$$\mathbf{e}^k = \sum_i a^{ki}\bar{\mathbf{e}}^i = \sum_i \bar{\mathbf{e}}^i \frac{\partial x^k}{\partial \bar{x}^i} \tag{31}$$

For the components of a vector represented covariantly, we have from the law of transformation (24),

$$\bar{v}_i = \sum_k a^{ki}v_k = \sum_k v_k \frac{\partial x^k}{\partial \bar{x}^i} \tag{32}$$

For the components of a vector represented contravariantly, we have from the law of transformation (22),

$$\bar{v}^i = \sum_k a_{ik}v^k = \sum_k v^k \frac{\partial \bar{x}^i}{\partial x^k} \tag{33}$$

If we have a general transformation of coordinates, say

$$\bar{x}^i = \bar{x}^i(x^1,x^2,x^3) \qquad i = 1, 2, 3$$

then any vector whose components transform according to the law (32) is called a **covariant vector,** and any vector whose components transform according to the law (33) is called a **contravariant vector.** In rectangular coordinates, as we pointed out earlier, the base set $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ and the reciprocal set $\mathbf{e}^1$, $\mathbf{e}^2$, $\mathbf{e}^3$ are identical Hence, there is no distinction between covariant and contravariant vectors, and no need to introduce the two concepts, in elementary vector analysis

EXERCISES

1 Prove that, for any nonsingular matrix A, the product $\bar{G} = (A^{-1})^T A^{-1}$ is symmetric

2 What is the condition that the set of base vectors $\bar{\mathbf{e}}_1$, $\bar{\mathbf{e}}_2$, $\bar{\mathbf{e}}_3$ and the set of reciprocal vectors $\bar{\mathbf{e}}^1$, $\bar{\mathbf{e}}^2$, $\bar{\mathbf{e}}^3$ be the same?

3 a Let x^1, x^2, x^3 and $\bar{x}^1$, $\bar{x}^2$, $\bar{x}^3$ be, respectively, rectangular and oblique coordinates connected by the transformation equations

$$\begin{aligned} \bar{x}^1 &= 2x^1 \qquad\quad + x^3 \\ \bar{x}^2 &= x^1 + 2x^2 + 3x^3 \\ \bar{x}^3 &= x^1 + x^2 + x^3 \end{aligned}$$

Working directly from their definitions, determine the rectangular representation of the base vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ and the reciprocal vectors $\mathbf{e}^1$, $\mathbf{e}^2$, $\mathbf{e}^3$. Thence verify that Eqs. (12) and (17) are satisfied.

b Work part **a** if the matrix of the transformation to oblique coordinates is

$$A = \begin{Vmatrix} 1 & 2 & -1 \\ 2 & 1 & 3 \\ 1 & 1 & 1 \end{Vmatrix}$$

4 **a** In Exercise 3a, what is the distance from the origin to the point whose oblique coordinates are (1,1,1)? What is the distance between the points whose oblique coordinates are (1,1,1) and (1,2,3)?

b In Exercise 3b what is the distance from the origin to the points whose oblique coordinates are (1,0,−1) and (2,1,1)?

5 **a** If $\bar{\mathbf{U}}$ and $\bar{\mathbf{V}}$ are two vectors represented contravariantly in an oblique coordinate system connected with a rectangular coordinate system by the transformation $\bar X = AX$, show that the angle between U and V is given by the formula

$$\cos\theta = \frac{\bar U^T \bar G \bar V}{\sqrt{\bar U^T \bar G \bar U}\,\sqrt{\bar V^T \bar G \bar V}}$$

b What is the angle between two vectors $\bar{\mathbf{U}}$ and $\bar{\mathbf{V}}$ represented covariantly?

13.3

Generalized coordinates

Let x^1, x^2, x^3 be three independent, single-valued, differentiable scalar point functions such that to every point of some region R of three-dimensional euclidean space there corresponds a unique triple of values (x^1,x^2,x^3), and such that to every triple of values (x^1,x^2,x^3) within ranges determined by the nature of R there corresponds a unique point of R. Then x^1, x^2, x^3 are called **generalized coordinates** in R, and the correspondence between the points of R and the number triples (x^1,x^2,x^3) is called a **generalized coordinate system** for R. Rectangular, cylindrical, spherical, and now oblique coordinates are familiar examples of generalized coordinates.

Through each point P of R there passes a unique surface S^1 on which x^1 is constant, a unique surface S^2 on which x^2 is constant, and a unique surface S^3 on which x^3 is constant. These surfaces intersect by pairs in curves, called **parametric curves,** which pass through P and on which one and only one of the generalized coordinates varies. Under the assumptions we have made about x^1, x^2, x^3, it can be shown that at each point of R the tangents to the parametric curves which pass through that point are noncoplanar. In general, the tangents to the parametric curves will vary in direction from point to point, and no one set of directions is singled out as any more natural than any other for the directions of a set of base vectors for R. However, *at each point*, vectors along the tangents to the parametric curves through that point provide a natural basis for the representation of vectors extending from

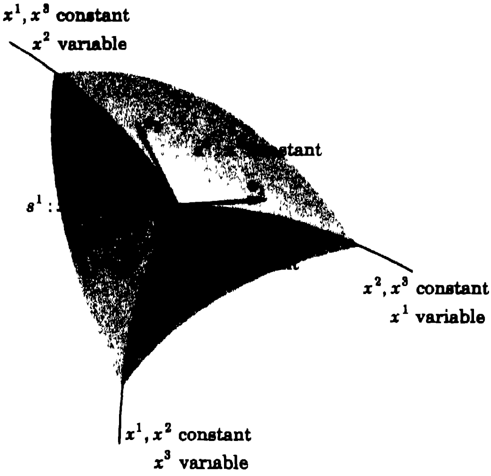


FIGURE 13.4 The parametric curves and the local base vectors at a point P in a generalized coordinate system

that point as origin (Fig. 13.4), and our development will be based on this concept of local base vectors and, of course, the related concept of local reciprocal base vectors

The **local base vectors** $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ at any point P we define to have, respectively, the directions of the tangents to the x^1-, x^2-, x^3-parametric curves at P, and to have lengths $|\mathbf{e}_i| = \sqrt{\mathbf{e}_i \cdot \mathbf{e}_i}$, such that, if ds is the infinitesimal distance along the x^i-parametric curve corresponding to the infinitesimal change dx^i in x^i, then

(1) $$ds = |\mathbf{e}_i\, dx^i| = \sqrt{\mathbf{e}_i \cdot \mathbf{e}_i}\, |dx^i|$$

At P we define the **local reciprocal base vectors** $\mathbf{e}^1$, $\mathbf{e}^2$, $\mathbf{e}^3$ precisely as we did in oblique coordinates, namely, by the conditions

$$\mathbf{e}^i \cdot \mathbf{e}_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

where, as usual, $\mathbf{e}^i \cdot \mathbf{e}_j = |\mathbf{e}^i|\, |\mathbf{e}_j| \cos(\mathbf{e}^i,\mathbf{e}_j)$.

Since our definitions for the local base vectors and the corresponding reciprocal vectors involve the notion of length, we must, of course, have some method of measuring distances To do this, we assume the existence throughout R of a positive-definite matrix

$$G = \|g_{ij}\|$$

whose elements are functions of the generalized coordinates and which has the property that, if

$$dX = \begin{Vmatrix} dx^1 \\ dx^2 \\ dx^3 \end{Vmatrix}$$

then the distance ds from P. (x^1,x^2,x^3) to Q: $(x^1 + dx^1, x^2 + dx^2, x^3 + dx^3)$ is given by the formula

(2) $$(ds)^2 = (dX)^T G(dX) = \sum_{ij} g_{ij}\, dx^i\, dx^j$$

Thus, if $x^1 = x^1(t)$, $x^2 = x^2(t)$, $x^3 = x^3(t)$ are the parametric equations of a curve, then the length of the curve between the points P_1 and P_2 at which t has the values t_1 and t_2, respectively, is

$$\int_{P_1}^{P_2} ds = \int_{P_1}^{P_2} \sqrt{\sum_{i,j} g_{ij}\, dx^i\, dx^j} = \int_{t_1}^{t_2} \sqrt{\sum_{i,j} g_{ij} \frac{dx^i}{dt}\frac{dx^j}{dt}}\, dt$$

In particular, for the length of an arbitrary infinitesimal vector

$$\mathbf{e}_1\, dx^1 + \mathbf{e}_2\, dx^2 + \mathbf{e}_3\, dx^3$$

we have

$$\begin{aligned}(ds)^2 &= (\mathbf{e}_1\, dx^1 + \mathbf{e}_2\, dx^2 + \mathbf{e}_3\, dx^3) \cdot (\mathbf{e}_1\, dx^1 + \mathbf{e}_2\, dx^2 + \mathbf{e}_3\, dx^3) \\ &= \sum_{i,j} \mathbf{e}_i \cdot \mathbf{e}_j\, dx^i\, dx^j = \sum_{i,j} g_{ij}\, dx^i\, dx^j\end{aligned}$$

Since the differentials of the coordinates are independent and arbitrary, the coefficients of corresponding terms in the last two sums must be identical, therefore,

(3) $$\mathbf{e}_i \cdot \mathbf{e}_j = g_{ij}$$

In particular,

(4) $$|\mathbf{e}_i| = \sqrt{\mathbf{e}_i \cdot \mathbf{e}_i} = \sqrt{g_{ii}}$$

Using (3) and (4), we can determine without integration the length of a noninfinitesimal vector $\mathbf{V} = v^1\mathbf{e}_1 + v^2\mathbf{e}_2 + v^3\mathbf{e}_3$ expressed in terms of the base vectors at a point P. In fact,

$$\begin{aligned}|\mathbf{V}|^2 = \mathbf{V} \cdot \mathbf{V} &= (v^1\mathbf{e}_1 + v^2\mathbf{e}_2 + v^3\mathbf{e}_3) \cdot (v^1\mathbf{e}_1 + v^2\mathbf{e}_2 + v^3\mathbf{e}_3) \\ &= \sum_{i,j} \mathbf{e}_i \cdot \mathbf{e}_j v^i v^j = \sum_{i,j} g_{ij} v^i v^j = V^T G V\end{aligned}$$

Hence,

(5) $$|\mathbf{V}| = \sqrt{V^T G V}$$

where, of course, the elements g_{ij} of G are to be evaluated at the point P at which $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ are the base vectors.

From (3) we also draw the important conclusion that *a necessary and sufficient condition that the parametric curves be orthogonal at every point of R is that $g_{ij} = 0$ for $i \neq j$ at all points of R*

By exactly the same reasoning we used to derive Eqs (7) and (10) in the last section we can now prove that, for the local base vectors and the local reciprocal base vectors, we have the following relations

(6) $$\mathbf{e}_i = \sum_k g_{ik}\mathbf{e}^k$$

(7) $$\mathbf{e}^i = \sum_k g^{ik}\mathbf{e}_k$$

where, as in the last section, g^{ik} is the element in the ith row and kth column of the matrix G^{-1} which is the inverse of $G = \|g_{ik}\|$. Furthermore, by forming the scalar product of Eq (7) with $\mathbf{e}^j$ and using the definitive relation $\mathbf{e}^j \cdot \mathbf{e}_k = \delta_k^j$, where δ_k^j is the

Kronecker delta,* we obtain the following companion result to Eq. (3):

(8) $$\mathbf{e}^i \cdot \mathbf{e}^j = g^{ij}$$

Equation (7) is not the only formula which can be used to express the local reciprocal vectors in terms of the local base vectors. Specifically, it is easy to verify that $\mathbf{e}^1$, $\mathbf{e}^2$, $\mathbf{e}^3$ are given in terms of $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ by the formulas

(9) $$\mathbf{e}^1 = \frac{\mathbf{e}_2 \times \mathbf{e}_3}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]} \qquad \mathbf{e}^2 = \frac{\mathbf{e}_3 \times \mathbf{e}_1}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]} \qquad \mathbf{e}^3 = \frac{\mathbf{e}_1 \times \mathbf{e}_2}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]}$$

Hence, using the result of Exercise 22, Sec. 12.1,

$$[\mathbf{e}^1\mathbf{e}^2\mathbf{e}^3] = \frac{\mathbf{e}_2 \times \mathbf{e}_3}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]} \cdot \frac{\mathbf{e}_3 \times \mathbf{e}_1}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]} \times \frac{\mathbf{e}_1 \times \mathbf{e}_2}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]} = \frac{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]^2}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]^3}$$

(10) $$= \frac{1}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]}$$

Moreover, using Eq. (6) in conjunction with Eq. (10), the numerical value of $[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]$ (and hence of $[\mathbf{e}^1\mathbf{e}^2\mathbf{e}^3]$) can easily be found. For, by (6),

$$[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3] = \Big(\sum_i g_{1i}\mathbf{e}^i\Big) \cdot \Big(\sum_j g_{2j}\mathbf{e}^j\Big) \times \Big(\sum_k g_{3k}\mathbf{e}^k\Big)$$

(11) $$= \sum_{i,j,k} g_{1i}g_{2j}g_{3k}[\mathbf{e}^i\mathbf{e}^j\mathbf{e}^k]$$

Now, of the $3^3 = 27$ terms which arise as i, j, and k range independently over the numbers 1, 2, 3, twenty-one are zero, because the scalar triple product $[\mathbf{e}^i\mathbf{e}^j\mathbf{e}^k]$ contains at least one repeated factor. Of the remaining six terms in the last sum there are three, corresponding to the sets of values $(i,j,k) = (1,2,3), (2,3,1), (3,1,2)$, in which

$$[\mathbf{e}^i\mathbf{e}^j\mathbf{e}^k] = [\mathbf{e}^1\mathbf{e}^2\mathbf{e}^3] = \frac{1}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]}$$

In the remaining three terms, corresponding to the sets of values $(i,j,k) = (1,3,2), (3,2,1), (2,1,3)$, the factor $[\mathbf{e}^i\mathbf{e}^j\mathbf{e}^k]$ is equal to

$$-[\mathbf{e}^1\mathbf{e}^2\mathbf{e}^3] = -\frac{1}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]}$$

Hence, factoring $1/[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]$ from the sum in (11) and cross-multiplying, we have

$$[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]^2 = g_{11}g_{22}g_{33} + g_{12}g_{23}g_{31} + g_{13}g_{21}g_{32} - g_{11}g_{23}g_{32} - g_{13}g_{22}g_{31} - g_{12}g_{21}g_{33}$$

Since the sum on the right in the last equation is precisely the expansion of the determinant of the matrix $G = \|g_{ij}\|$, we have thus established the useful result

(12) $$[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]^2 = |G|$$

---

* Defined at the end of Sec. 10.1

where G is the matrix which defines the metrical properties of space.

We now turn our attention to transformations from one set of generalized coordinates to another. In particular, we are interested in the laws of transformation for the fundamental matrices G and G^{-1}, the local base vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, the local reciprocal vectors $\mathbf{e}^1$, $\mathbf{e}^2$, $\mathbf{e}^3$, and vectors expressed in terms of these reference vectors, which are induced by a transformation of coordinates.

Let us suppose, then, that we have two systems of coordinates (x^1,x^2,x^3) and $(\bar{x}^1,\bar{x}^2,\bar{x}^3)$ connected by transformation equations of the form

$$\begin{aligned}\bar{x}^1 &= \bar{x}^1(x^1,x^2,x^3)\\ \bar{x}^2 &= \bar{x}^2(x^1,x^2,x^3)\\ \bar{x}^3 &= \bar{x}^3(x^1,x^2,x^3)\end{aligned}$$

or, simply,

(13) $$T \qquad \bar{x}^i = \bar{x}^i(x^1,x^2,x^3) \qquad i = 1, 2, 3$$

In particular cases, the equations (13) might be the equations connecting a rectangular and an oblique coordinate system, as in the last section, a rectangular and a cylindrical coordinate system; a rectangular and a spherical coordinate system, or a cylindrical and a spherical coordinate system

Naturally, we wish a point with coordinates (x^1,x^2,x^3) in the x-system to have a unique set of coordinates $(\bar{x}^1,\bar{x}^2,\bar{x}^3)$ in the $\bar{x}$-system Hence we require that, throughout the region R with which we are concerned, the $\bar{x}^i$'s be single-valued functions of the x^i's. Moreover, we wish the point with coordinates $(\bar{x}^1,\bar{x}^2,\bar{x}^3)$ to have a unique set of x-coordinates. Hence, we also require that Eqs. (13) be solvable for x^1, x^2, x^3 as single-valued functions of $\bar{x}^1$, $\bar{x}^2$, $\bar{x}^3$, say

(14) $$T^{-1} \qquad x^i = x^i(\bar{x}^1,\bar{x}^2,\bar{x}^3) \qquad i = 1, 2, 3$$

In advanced calculus it is shown* that, if the first partial derivatives of the coordinate functions $\bar{x}^i$ in T are continuous and if, throughout R, the so-called **Jacobian determinant**

(15) $$|J| = \left|\frac{\partial(\bar{x}^1,\bar{x}^2,\bar{x}^3)}{\partial(x^1,x^2,x^3)}\right|^{\dagger} = \begin{vmatrix}\dfrac{\partial\bar{x}^1}{\partial x^1} & \dfrac{\partial\bar{x}^1}{\partial x^2} & \dfrac{\partial\bar{x}^1}{\partial x^3}\\ \dfrac{\partial\bar{x}^2}{\partial x^1} & \dfrac{\partial\bar{x}^2}{\partial x^2} & \dfrac{\partial\bar{x}^2}{\partial x^3}\\ \dfrac{\partial\bar{x}^3}{\partial x^1} & \dfrac{\partial\bar{x}^3}{\partial x^2} & \dfrac{\partial\bar{x}^3}{\partial x^3}\end{vmatrix}$$

---

* See, for instance, R C. Buck, "Advanced Calculus," p 215, McGraw-Hill Book Company, New York, 1956.

† Named for the German mathematician C G J Jacobi (1804-1851) We shall frequently refer to the **Jacobian matrix** J simply as the **Jacobian** of the transformation T

is different from zero, as we shall suppose, then around any interior point of R there exists a neighborhood in which T has a a single-valued inverse (14) Naturally, since the equations of the inverse transformation (14) are to be uniquely solvable for x^1, x^2, x^3, the Jacobian determinant of the inverse transformation

$$|\bar{J}| = \left|\frac{\partial(x^1,x^2,x^3)}{\partial(\bar{x}^1,\bar{x}^2,\bar{x}^3)}\right| = \begin{vmatrix} \dfrac{\partial x^1}{\partial \bar{x}^1} & \dfrac{\partial x^1}{\partial \bar{x}^2} & \dfrac{\partial x^1}{\partial \bar{x}^3} \\ \dfrac{\partial x^2}{\partial \bar{x}^1} & \dfrac{\partial x^2}{\partial \bar{x}^2} & \dfrac{\partial x^2}{\partial \bar{x}^3} \\ \dfrac{\partial x^3}{\partial \bar{x}^1} & \dfrac{\partial x^3}{\partial \bar{x}^2} & \dfrac{\partial x^3}{\partial \bar{x}^3} \end{vmatrix} \tag{16}$$

must also be different from zero throughout R.

We have now reached the point where it is convenient, or indeed necessary, to introduce the so-called **Einstein summation convention.** Just as the summation symbol Σ effects a great notational economy when it is used instead of writing a sum of terms at length, so this summation convention replaces the symbol Σ with a notation still shorter and much more suggestive. Briefly, the convention is this: *If any term contains the same letter twice as a distinguishing index, it is understood that the term is to be summed for all values of the repeated index* For example, using the summation convention, with the understanding that the range of our indices is 1 to 3, we can write the differential of $\bar{x}^i$ in the equivalent forms

$$d\bar{x}^i = \frac{\partial \bar{x}^i}{\partial x^1}dx^1 + \frac{\partial \bar{x}^i}{\partial x^2}dx^2 + \frac{\partial \bar{x}^i}{\partial x^3}dx^3 = \sum_{j=1}^{3}\frac{\partial \bar{x}^i}{\partial x^j}dx^j = \frac{\partial \bar{x}^i}{\partial x^j}dx^j$$

In the last expression, the index i identifies the particular variable $\bar{x}^i$ whose differential is being considered, and cannot be changed On the other hand, the index j merely indicates that summation over a certain range is to be carried out, and, like the variable of integration in a definite integral, can be changed at pleasure to any other letter except i, of course. Thus we can write equally well

$$d\bar{x}^i = \frac{\partial \bar{x}^i}{\partial x^j}dx^j = \frac{\partial \bar{x}^i}{\partial x^k}dx^k = \frac{\partial \bar{x}^i}{\partial x^\alpha}dx^\alpha = \cdot$$

An index which can thus be arbitrarily replaced by another is usually called a **dummy index** or an **umbral index.**

The summation convention also permits more than one pair of repeated indices in a term to be summed For instance, applying the convention first to the repeated index i and then to j, we have

$$\begin{aligned} g_{ij}\,dx^i\,dx^j &= g_{1j}\,dx^1\,dx^j + g_{2j}\,dx^2\,dx^j + g_{3j}\,dx^3\,dx^j \\ &= (g_{11}\,dx^1\,dx^1 + g_{12}\,dx^1\,dx^2 + g_{13}\,dx^1\,dx^3) \\ &\quad + (g_{21}\,dx^2\,dx^1 + g_{22}\,dx^2\,dx^2 + g_{23}\,dx^2\,dx^3) \\ &\quad + (g_{31}\,dx^3\,dx^1 + g_{32}\,dx^3\,dx^2 + g_{33}\,dx^3\,dx^3) \\ &= \sum_{i,j} g_{ij}\,dx^i\,dx^j \end{aligned}$$

It should be noted that

$$g_{ij}\,dx^i\,dx^j \neq g_{ii}\,dx^i\,dx^i$$

since the latter is equal to the simpler sum

$$g_{11}\,dx^1\,dx^1 + g_{22}\,dx^2\,dx^2 + g_{33}\,dx^3\,dx^3$$

Hence, unless the more restricted meaning is intended, the same index cannot be used a second time in the same term as a dummy index.

Preparatory to resuming our discussion of coordinate transformations, it will be helpful to introduce several simple lemmas at this point.

LEMMA 1

If (x^1,x^2,x^3) and $(\bar{x}^1,\bar{x}^2,\bar{x}^3)$ are coordinates connected by a transformation

$$\bar{x}^i = \bar{x}^i(x^1,x^2,x^3)$$

then $$\frac{\partial \bar{x}^i}{\partial x^\alpha}\frac{\partial x^\alpha}{\partial \bar{x}^j} = \delta_j^i$$

PROOF By hypothesis, $\bar{x}^i$ is a differentiable function of x^1, x^2, x^3, which in turn are differentiable functions of $\bar{x}^1$, $\bar{x}^2$, $\bar{x}^3$ Hence, by the chain rule of partial differentiation,

$$\frac{\partial \bar{x}^i}{\partial \bar{x}^i} = 1 = \frac{\partial \bar{x}^i}{\partial x^1}\frac{\partial x^1}{\partial \bar{x}^i} + \frac{\partial \bar{x}^i}{\partial x^2}\frac{\partial x^2}{\partial \bar{x}^i} + \frac{\partial \bar{x}^i}{\partial x^3}\frac{\partial x^3}{\partial \bar{x}^i} = \frac{\partial \bar{x}^i}{\partial x^\alpha}\frac{\partial x^\alpha}{\partial \bar{x}^i} \qquad i \text{ not summed}$$

and, since the $\bar{x}^i$'s are independent,

$$\frac{\partial \bar{x}^i}{\partial \bar{x}^j} = 0 = \frac{\partial \bar{x}^i}{\partial x^1}\frac{\partial x^1}{\partial \bar{x}^j} + \frac{\partial \bar{x}^i}{\partial x^2}\frac{\partial x^2}{\partial \bar{x}^j} + \frac{\partial \bar{x}^i}{\partial x^3}\frac{\partial x^3}{\partial \bar{x}^j} = \frac{\partial \bar{x}^i}{\partial x^\alpha}\frac{\partial x^\alpha}{\partial \bar{x}^j} \qquad i \neq j$$

These two relations together establish the assertion of the lemma Of course, by an identical proof it follows that

$$\frac{\partial x^i}{\partial \bar{x}^\alpha}\frac{\partial \bar{x}^\alpha}{\partial x^j} = \delta_j^i$$

LEMMA 2

If ϕ^i and $\bar{\phi}^i$ $(i = 1, 2, 3)$ are, respectively, functions of x^1, x^2, x^3 and $\bar{x}^1$, $\bar{x}^2$, $\bar{x}^3$, then

$$\bar{\phi}^i = \phi^\alpha\frac{\partial \bar{x}^i}{\partial x^\alpha} \qquad \text{implies} \qquad \bar{\phi}^i\frac{\partial x^\alpha}{\partial \bar{x}^i} = \phi^\alpha$$

and conversely.

PROOF To provide us with further insight into the efficiency of the summation convention, let us first prove this lemma using the more familiar Σ notation We are given the relation

$$\bar{\phi}^i = \sum_{\alpha=1}^{3}\phi^\alpha\frac{\partial \bar{x}^i}{\partial x^\alpha} = \sum_{\beta=1}^{3}\phi^\beta\frac{\partial \bar{x}^i}{\partial x^\beta}$$

If we now multiply both sides of this equation in its second form by $\dfrac{\partial x^\alpha}{\partial \bar{x}^i}$ and then sum over i, we have

$$\sum_{i=1}^{3} \bar{\phi}^i \frac{\partial x^\alpha}{\partial \bar{x}^i} = \sum_{i=1}^{3} \left(\frac{\partial x^\alpha}{\partial \bar{x}^i} \sum_{\beta=1}^{3} \phi^\beta \frac{\partial \bar{x}^i}{\partial x^\beta} \right)$$

or, interchanging the order of summation on the right,

$$\sum_{i=1}^{3} \bar{\phi}^i \frac{\partial x^\alpha}{\partial \bar{x}^i} = \sum_{\beta=1}^{3} \left(\phi^\beta \sum_{i=1}^{3} \frac{\partial x^\alpha}{\partial \bar{x}^i} \frac{\partial \bar{x}^i}{\partial x^\beta} \right)$$

Now, by Lemma 1, the inner sum on the right is equal to $\delta_\beta{}^\alpha$ Hence, the right-hand side reduces to

$$\sum_{\beta=1}^{3} \delta_\beta{}^\alpha \phi^\beta$$

which is equal to zero unless $\beta = \alpha$ Therefore, finally,

$$\sum_{\alpha=1}^{3} \bar{\phi}^i \frac{\partial x^\alpha}{\partial \bar{x}^i} \equiv \bar{\phi}^i \frac{\partial x^\alpha}{\partial \bar{x}^i} = \phi^\alpha \qquad \text{as asserted}$$

Using the summation convention, our proof would have proceeded as follows Introducing the dummy index β in place of α, we begin with

$$\bar{\phi}^i = \phi^\beta \frac{\partial \bar{x}^i}{\partial x^\beta}$$

Now, multiplying both sides by $\dfrac{\partial x^\alpha}{\partial \bar{x}^i}$ and using Lemma 1, we have

$$\bar{\phi}^i \frac{\partial x^\alpha}{\partial \bar{x}^i} = \phi^\beta \frac{\partial x^\alpha}{\partial \bar{x}^i} \frac{\partial \bar{x}^i}{\partial x^\beta} = \phi^\beta \delta_\beta{}^\alpha = \phi^\alpha$$

The converse assertion is, of course, established in exactly the same fashion

LEMMA 3

If ϕ^{ij} and $\bar{\phi}^{ij}$ $(i,j = 1, 2, 3)$ are, respectively, functions of x^1, x^2, x^3 and $\bar{x}^1$, $\bar{x}^2$, $\bar{x}^3$, then any one of the relations

$$\bar{\phi}^{ij} = \phi^{\alpha\beta} \frac{\partial \bar{x}^i}{\partial x^\alpha} \frac{\partial \bar{x}^j}{\partial x^\beta}$$

$$\bar{\phi}^{ij} \frac{\partial x^\beta}{\partial \bar{x}^j} = \phi^{\alpha\beta} \frac{\partial \bar{x}^i}{\partial x^\alpha}$$

$$\bar{\phi}^{ij} \frac{\partial x^\alpha}{\partial \bar{x}^i} = \phi^{\alpha\beta} \frac{\partial \bar{x}^j}{\partial x^\beta}$$

$$\bar{\phi}^{ij} \frac{\partial x^\alpha}{\partial \bar{x}^i} \frac{\partial x^\beta}{\partial \bar{x}^j} = \phi^{\alpha\beta}$$

implies each of the others.

PROOF Because of the near-identity of the arguments, it will be sufficient to establish just one of the assertions of the lemma, say the assertion that

$$\bar{\phi}^{ij} = \phi^{\alpha\beta} \frac{\partial \bar{x}^i}{\partial x^\alpha} \frac{\partial \bar{x}^j}{\partial x^\beta} \qquad \text{implies} \qquad \bar{\phi}^{ij} \frac{\partial x^\alpha}{\partial \bar{x}^i} \frac{\partial x^\beta}{\partial \bar{x}^j} = \phi^{\alpha\beta}$$

To do this, let us write the first relation using a and b in place of α and β, and then let us multiply both sides by $\frac{\partial x^\alpha}{\partial \bar{x}^i}\frac{\partial x^\beta}{\partial \bar{x}^j}$ and sum over i and j. This gives us, by Lemma 1,

$$\begin{aligned}\bar{\phi}^{ij}\frac{\partial x^\alpha}{\partial \bar{x}^i}\frac{\partial x^\beta}{\partial \bar{x}^j} &= \phi^{ab}\left(\frac{\partial x^\alpha}{\partial \bar{x}^i}\frac{\partial \bar{x}^i}{\partial x^a}\right)\left(\frac{\partial x^\beta}{\partial \bar{x}^j}\frac{\partial \bar{x}^j}{\partial x^b}\right)\\ &= \phi^{ab}\,\delta_a^{\alpha}\,\delta_b^{\beta}\\ &= \phi^{\alpha\beta} \qquad \text{as asserted.}\end{aligned}$$

In Secs. 10.2 and 10.3, when we considered linear transformations such as

$$T_1 \quad Y = AX \qquad \text{and} \qquad T_2 \quad Z = BY \qquad A,B \text{ nonsingular}$$

we observed that the matrices of the inverse transformations T_1^{-1} and T_2^{-1} are A^{-1} and B^{-1}, respectively, and that the matrix of the transformation resulting when T_1 is followed by T_2 is BA. Since linear transformations are obviously special cases of the transformation (13), it is natural to ask whether general coordinate transformations have comparable properties. The answer is Yes, and in fact we have the following theorems, the proof of the second of which we shall leave as an exercise.

THEOREM 1

If T: $\bar{x}^\alpha = \bar{x}^\alpha(x^1,x^2,x^3)$ is a transformation with Jacobian J, then the Jacobian $\bar{J}$ of the inverse transformation T^{-1}: $x^\alpha = x^\alpha(\bar{x}^1,\bar{x}^2,\bar{x}^3)$ is J^{-1}.

PROOF By definition, the Jacobian of the direct transformation T is $J = \left\|\frac{\partial \bar{x}^i}{\partial x^k}\right\|$, and the Jacobian of the inverse transformation T^{-1} is $\bar{J} = \left\|\frac{\partial x^k}{\partial \bar{x}^j}\right\|$. From the definition of matrix multiplication, the element in the ith row and jth column of the product $J\bar{J}$ is $\frac{\partial \bar{x}^i}{\partial x^k}\frac{\partial x^k}{\partial \bar{x}^j}$, and, by Lemma 1, this sum is equal to δ_j^i. Thus,

$$J\bar{J} = \|\delta_j^i\| = I$$

Hence, $\bar{J} = J^{-1}$, that is, the matrix $\bar{J}$ of the inverse transformation T^{-1} is the inverse of the matrix J of the direct transformation T, as asserted.

COROLLARY 1

If J is the Jacobian of the transformation T: $\bar{x}^\alpha = \bar{x}^\alpha(x^1,x^2,x^3)$, then $\frac{\partial x^i}{\partial \bar{x}^j}$ is equal to $1/|J|$ times the cofactor of $\frac{\partial \bar{x}^j}{\partial x^i}$ in $|J|$.

THEOREM 2

If T_1: $\bar{x}^\alpha = \bar{x}^\alpha(x^1,x^2,x^3)$ is a transformation with Jacobian J_1 and if

$$T_2 \qquad \bar{\bar{x}}^\beta = \bar{\bar{x}}^\beta(\bar{x}^1,\bar{x}^2,\bar{x}^3)$$

is a transformation with Jacobian J_2, then the Jacobian of the transformation T_2T_1 is J_2J_1.

Let us now determine how the fundamental differential quadratic form $(ds)^2 = g_{ij}\,dx^i\,dx^j$ transforms when the coordinates x^1, x^2, x^3 are transformed into the coordinates $\bar{x}^1, \bar{x}^2, \bar{x}^3$ by means of Eqs. (13) For dx^i and dx^j we have, of course,

$$dx^i = \frac{\partial x^i}{\partial \bar{x}^\alpha}\,d\bar{x}^\alpha \qquad \text{and} \qquad dx^j = \frac{\partial x^j}{\partial \bar{x}^\beta}\,d\bar{x}^\beta$$

Hence, $(ds)^2$ becomes the quadratic form

$$g_{ij}\frac{\partial x^i}{\partial \bar{x}^\alpha}\frac{\partial x^j}{\partial \bar{x}^\beta}\,d\bar{x}^\alpha\,d\bar{x}^\beta$$

Therefore, if we write the quadratic form after transformation as

$$\bar{g}_{\alpha\beta}\,d\bar{x}^\alpha\,d\bar{x}^\beta$$

it follows that the coefficients $\bar{g}_{\alpha\beta}$ transform according to the law

$$\bar{g}_{\alpha\beta} = g_{ij}\frac{\partial x^i}{\partial \bar{x}^\alpha}\frac{\partial x^j}{\partial \bar{x}^\beta} \tag{17}$$

as we verified in the particular case of a transformation from rectangular coordinates to oblique coordinates in the last section [Eq (26)] Of course, considering the transformation from $\bar{x}$-coordinates to x-coordinates, an argument identical with the one we have just given provides us with the companion formula

$$g_{ij} = \bar{g}_{\alpha\beta}\frac{\partial \bar{x}^\alpha}{\partial x^i}\frac{\partial \bar{x}^\beta}{\partial x^j} \tag{18}$$

which also follows from an obvious modification of Lemma 3

Formula (18) also leads to the following interesting conclusion. From the rule for the multiplication of matrices, the element in the ith row and jth column of the product $J^T\bar{G}J$ is

$$\frac{\partial \bar{x}^\alpha}{\partial x^i}\,\bar{g}_{\alpha\beta}\,\frac{\partial \bar{x}^\beta}{\partial x^j}$$

However, by (18), this is the element g_{ij} in the ith row and jth column of G. Hence, taking determinants,

$$\left|\frac{\partial \bar{x}^\alpha}{\partial x^i}\,\bar{g}_{\alpha\beta}\,\frac{\partial \bar{x}^\beta}{\partial x^j}\right| = |g_{ij}| \qquad \text{or} \qquad |J^T\bar{G}J| = |G|$$

or, finally,

$$|J|^2|\bar{G}| = |G| \tag{19}$$

EXAMPLE 1

Obtain the formula for the differential of arc length in spherical coordinates

Since we know the formula for the differential of arc length in rectangular coordinates, namely,

$$(ds)^2 = (dx)^2 + (dy)^2 + (dz)^2 \tag{20}$$

we can obtain the corresponding formula in spherical coordinates by transformation from rectangular coordinates To do this, let x^1, x^2, x^3 denote, respectively, the rectangular coordinates x, y, z, and let $\bar{x}^1, \bar{x}^2, \bar{x}^3$ denote, respectively, the spherical coordinates r, θ, ϕ (Fig. 13 5). Then, as

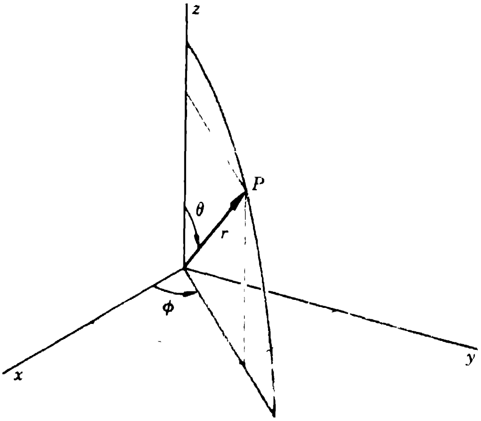


FIGURE 13 5
Plot showing the relation between rectangular and spherical coordinates

usual, we have

$$T^{-1} \qquad \begin{aligned} x^1 &= \bar{x}^1 \sin \bar{x}^2 \cos \bar{x}^3 \\ x^2 &= \bar{x}^1 \sin \bar{x}^2 \sin \bar{x}^3 \\ x^3 &= \bar{x}^1 \cos \bar{x}^2 \end{aligned}$$

and from these

$$\frac{\partial x^1}{\partial \bar{x}^1} = \sin \bar{x}^2 \cos \bar{x}^3 \qquad \frac{\partial x^1}{\partial \bar{x}^2} = \bar{x}^1 \cos \bar{x}^2 \cos \bar{x}^3 \qquad \frac{\partial x^1}{\partial \bar{x}^3} = -\bar{x}^1 \sin \bar{x}^2 \sin \bar{x}^3$$

$$\frac{\partial x^2}{\partial \bar{x}^1} = \sin \bar{x}^2 \sin \bar{x}^3 \qquad \frac{\partial x^2}{\partial \bar{x}^2} = \bar{x}^1 \cos \bar{x}^2 \sin \bar{x}^3 \qquad \frac{\partial x^2}{\partial \bar{x}^3} = \bar{x}^1 \sin \bar{x}^2 \cos \bar{x}^3$$

$$\frac{\partial x^3}{\partial \bar{x}^1} = \cos \bar{x}^2 \qquad \frac{\partial x^3}{\partial \bar{x}^2} = -\bar{x}^1 \sin \bar{x}^2 \qquad \frac{\partial x^3}{\partial \bar{x}^3} = 0$$

Hence, substituting into Eq (17) and noting from Eq (20) that $g_{ij} = \delta_i^j$, we have

$$\begin{aligned} g_{11} &= (\sin \bar{x}^2 \cos \bar{x}^3)^2 + (\sin \bar{x}^2 \sin \bar{x}^3)^2 + (\cos \bar{x}^2)^2 \\ &= 1 \\ \bar{g}_{22} &= (\bar{x}^1 \cos \bar{x}^2 \cos \bar{x}^3)^2 + (\bar{x}^1 \cos \bar{x}^2 \sin \bar{x}^3)^2 + (-\bar{x}^1 \cos \bar{x}^2)^2 \\ &= (\bar{x}^1)^2 \\ \bar{g}_{33} &= (-\bar{x}^1 \sin \bar{x}^2 \sin \bar{x}^3)^2 + (\bar{x}^1 \sin \bar{x}^2 \cos \bar{x}^3)^2 \\ &= (\bar{x}^1 \sin \bar{x}^2)^2 \\ \bar{g}_{12} &= (\sin \bar{x}^2 \cos \bar{x}^3)(\bar{x}^1 \cos \bar{x}^2 \cos \bar{x}^3) + (\sin \bar{x}^2 \sin \bar{x}^3)(\bar{x}^1 \cos \bar{x}^2 \sin \bar{x}^3) + (\cos \bar{x}^2)(-\bar{x}^1 \sin \bar{x}^2) \\ &= 0 \\ \bar{g}_{13} &= (\sin \bar{x}^2 \cos \bar{x}^3)(-\bar{x}^1 \sin \bar{x}^2 \sin \bar{x}^3) + (\sin \bar{x}^2 \sin \bar{x}^3)(\bar{x}^1 \sin \bar{x}^2 \cos \bar{x}^3) \\ &= 0 \\ \bar{g}_{23} &= (\bar{x}^1 \cos \bar{x}^2 \cos \bar{x}^3)(-\bar{x}^1 \sin \bar{x}^2 \sin \bar{x}^3) + (\bar{x}^1 \cos \bar{x}^2 \sin \bar{x}^3)(\bar{x}^1 \sin \bar{x}^2 \cos \bar{x}^3) \\ &= 0 \end{aligned}$$

and, finally,

$$\begin{aligned} (ds)^2 &= \bar{g}_{ij}\, d\bar{x}^i\, d\bar{x}^j \\ &= (d\bar{x}^1)^2 + (\bar{x}^1)^2 (d\bar{x}^2)^2 + (\bar{x}^1 \sin \bar{x}^2)^2 (d\bar{x}^3)^2 \\ &= (dr)^2 + r^2 (d\theta)^2 + (r \sin \theta)^2 (d\phi)^2 \end{aligned}$$

When the coordinates x^1, x^2, x^3 are replaced by the coordinates $\bar{x}^1, \bar{x}^2, \bar{x}^3$, there is, of course, a new set of parametric curves passing through an arbitrary point P and, hence, a new set of base vectors $\bar{\mathbf{e}}_1, \bar{\mathbf{e}}_2, \bar{\mathbf{e}}_3$ and a new set of reciprocal base vectors $\bar{\mathbf{e}}^1, \bar{\mathbf{e}}^2, \bar{\mathbf{e}}^3$ To obtain the relations between the old and the new base vectors, let us consider an arbitrary infinitesimal displacement $d\mathbf{s}$ expressed in terms of each system.

$$d\mathbf{s} = dx^\alpha\, \mathbf{e}_\alpha = d\bar{x}^i\, \bar{\mathbf{e}}_i \tag{21}$$

Now, from the transformation equations, we have

$$d\bar{x}^i = \frac{\partial \bar{x}^i}{\partial x^\alpha}\, dx^\alpha$$

and, hence, from (21), we can write

$$dx^\alpha\, \mathbf{e}_\alpha = \frac{\partial \bar{x}^i}{\partial x^\alpha}\, dx^\alpha\, \bar{\mathbf{e}}_i$$

Now the differentials are arbitrary, therefore, the coefficients of corresponding differentials on each side of the last equation must be equal. Thus,

$$\mathbf{e}_\alpha = \frac{\partial \bar{x}^i}{\partial x^\alpha}\, \bar{\mathbf{e}}_i \tag{22}$$

Similarly, or by Lemma 2,

$$\bar{\mathbf{e}}_i = \frac{\partial x^\alpha}{\partial \bar{x}^i}\, \mathbf{e}_\alpha \tag{23}$$

Formulas (29) and (28) of Sec. 13 2 were, respectively, of course, special cases of these relations

Knowing from Eq. (6) how the local base vectors are expressed in terms of the local reciprocal base vectors in any coordinate system and knowing from Eq. (23) how the local base vectors transform, we can now determine how the local reciprocal base vectors transform For, beginning with the relation (6) for the new coordinate system, namely,

$$\bar{\mathbf{e}}_i = \bar{g}_{ij}\bar{\mathbf{e}}^j$$

and substituting from Eqs (17) and (23), we have

$$\frac{\partial x^\alpha}{\partial \bar{x}^i}\, \mathbf{e}_\alpha = g_{\alpha\beta}\frac{\partial x^\alpha}{\partial \bar{x}^i}\frac{\partial x^\beta}{\partial \bar{x}^j}\, \bar{\mathbf{e}}^j$$

Now, multiplying this equation by $\dfrac{\partial \bar{x}^i}{\partial x^\gamma}$ and summing each side over i, we obtain

$$\left(\frac{\partial x^\alpha}{\partial \bar{x}^i}\frac{\partial \bar{x}^i}{\partial x^\gamma}\right)\mathbf{e}_\alpha = g_{\alpha\beta}\left(\frac{\partial x^\alpha}{\partial \bar{x}^i}\frac{\partial \bar{x}^i}{\partial x^\gamma}\right)\frac{\partial x^\beta}{\partial \bar{x}^j}\, \bar{\mathbf{e}}^j$$

or, using Lemma 1,

$$\delta_\gamma{}^\alpha \mathbf{e}_\alpha = g_{\alpha\beta}\, \delta_\gamma{}^\alpha \frac{\partial x^\beta}{\partial \bar{x}^j}\, \bar{\mathbf{e}}^j \qquad \text{and} \qquad \mathbf{e}_\gamma = g_{\gamma\beta}\frac{\partial x^\beta}{\partial \bar{x}^j}\, \bar{\mathbf{e}}^j$$

Now, if we multiply the last equation by $g^{\lambda\gamma}$ and sum over γ, making use of the fact that $\|g^{ij}\|$ is the inverse of $\|g_{ij}\|$ and, hence, that $g^{\lambda\gamma}g_{\gamma\beta} = \delta_\beta{}^\lambda$, we have

$$g^{\lambda\gamma}\mathbf{e}_\gamma = g^{\lambda\gamma}g_{\gamma\beta}\frac{\partial x^\beta}{\partial \bar{x}^j}\bar{\mathbf{e}}^j = \delta_\beta{}^\lambda\frac{\partial x^\beta}{\partial \bar{x}^j}\bar{\mathbf{e}}^j$$

and, using Eq. (7),

(24) $$\mathbf{e}^\lambda = \bar{\mathbf{e}}^j\frac{\partial x^\lambda}{\partial \bar{x}^j}$$

Similarly, or by using Lemma 2,

(25) $$\bar{\mathbf{e}}^j = \mathbf{e}^\lambda\frac{\partial \bar{x}^j}{\partial x^\lambda}$$

Equations (31) and (30) of Sec. 13.2 were, of course, respectively, special cases of these relations.

From Eq. (8) applied to the new coordinate system, we have

$$\bar{\mathbf{e}}^i\cdot\bar{\mathbf{e}}^j = \bar{g}^{ij}$$

Hence, using (25), we have

$$\left(\mathbf{e}^\alpha\frac{\partial \bar{x}^i}{\partial x^\alpha}\right)\cdot\left(\mathbf{e}^\beta\frac{\partial \bar{x}^j}{\partial x^\beta}\right) = \bar{g}^{ij}$$

or, since $\mathbf{e}^\alpha\cdot\mathbf{e}^\beta = g^{\alpha\beta}$,

(26) $$\bar{g}^{ij} = g^{\alpha\beta}\frac{\partial \bar{x}^i}{\partial x^\alpha}\frac{\partial \bar{x}^j}{\partial x^\beta}$$

which is the law of transformation for the g^{ij}'s. Equation (27), Sec. 13.2, is a special case of this result. Similarly, or by Lemma 3,

$$g^{\alpha\beta} = \bar{g}^{ij}\frac{\partial x^\alpha}{\partial \bar{x}^i}\frac{\partial x^\beta}{\partial \bar{x}^j}$$

When a vector represented contravariantly, that is, a vector expressed in terms of the local base vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, say

$$V^\alpha = v^1\mathbf{e}_1 + v^2\mathbf{e}_2 + v^3\mathbf{e}_3 = v^\alpha\mathbf{e}_\alpha$$

is expressed in terms of the corresponding local base vectors $\bar{\mathbf{e}}_1$, $\bar{\mathbf{e}}_2$, $\bar{\mathbf{e}}_3$ of a new coordinate system, we have, using (22), the new representation

$$\bar{V}^\alpha = v^\alpha\left(\frac{\partial \bar{x}^i}{\partial x^\alpha}\bar{\mathbf{e}}_i\right) = \left(\frac{\partial \bar{x}^i}{\partial x^\alpha}v^\alpha\right)\bar{\mathbf{e}}_i = \bar{v}^i\bar{\mathbf{e}}_i$$

Hence, the components of a contravariant vector transform according to the law

(27) $$\bar{v}^i = \frac{\partial \bar{x}^i}{\partial x^\alpha}v^\alpha$$

Similarly, for a vector represented covariantly, that is, a vector expressed in terms of the local reciprocal base vectors, say

$$V_\alpha = v_1\mathbf{e}^1 + v_2\mathbf{e}^2 + v_3\mathbf{e}^3 = v_\alpha\mathbf{e}^\alpha$$

we have, using (24), the new representation

$$\bar{V}_\alpha = v_\alpha\left(\bar{\mathbf{e}}^i \frac{\partial x^\alpha}{\partial \bar{x}^i}\right) = \left(v_\alpha \frac{\partial x^\alpha}{\partial \bar{x}^i}\right)\bar{\mathbf{e}}^i = \bar{v}_i\bar{\mathbf{e}}^i$$

Hence, the components of a covariant vector transform according to the law

$$\bar{v}_i = v_\alpha \frac{\partial x^\alpha}{\partial \bar{x}^i} \tag{28}$$

Equations (33) and (32), Sec. 13.2, were special cases of Eqs (27) and (28), respectively

EXERCISES

1 If the range of each index is 3, write out each of the following sums

a $f(x_i)\,\Delta x_i$ **b** $a_{ij}x_ix_j$

c $a_{ij}x_iy_j$ **d** $a_{ii}x_ix_i$

e $(a_ix^i)^2$ **f** $\dfrac{\partial x^i}{\partial y^j}\dfrac{\partial y^j}{\partial x^k}z_i$

g $\dfrac{\partial x^i}{\partial y^j}\dfrac{\partial y^j}{\partial x^k}z^k$

2 If the range of each index is n, show that

a $\delta_j^i\,\delta_k^j = \delta_k^i$ **b** $\delta_i^i = \delta_j^j = n$

c $\delta_j^kA_j = A_k$ **d** $\delta_k^iA^iA_k = A^kA_k$

e $\delta_j^iA_{ijk}\,\delta_l^k = A_{iil}$

3 Write out the proofs of Theorems 1 and 6, Sec 10 2, using the summation convention rather than the Σ notation

4 Write out the proofs of the remaining assertions of Lemma 3

5 Prove the following lemma If ϕ_{ij} and $\bar{\phi}_{ij}$ $(i,j = 1, 2, 3)$ are, respectively, functions of x^1, x^2, x^3 and $\bar{x}^1$, $\bar{x}^2$, $\bar{x}^3$, then any one of the relations

$$\bar{\phi}_{ij} = \phi_{\alpha\beta}\frac{\partial x^\alpha}{\partial \bar{x}^i}\frac{\partial x^\beta}{\partial \bar{x}^j}$$

$$\bar{\phi}_{ij}\frac{\partial \bar{x}^j}{\partial x^\beta} = \phi_{\alpha\beta}\frac{\partial x^\alpha}{\partial \bar{x}^i}$$

$$\bar{\phi}_{ij}\frac{\partial \bar{x}^i}{\partial x^\alpha} = \phi_{\alpha\beta}\frac{\partial x^\beta}{\partial \bar{x}^j}$$

$$\bar{\phi}_{ij}\frac{\partial \bar{x}^i}{\partial x^\alpha}\frac{\partial \bar{x}^j}{\partial x^\beta} = \phi_{\alpha\beta}$$

implies each of the others

Each of the following problems refers to a cylindrical coordinate system

6 **a** What is the differential of arc in cylindrical coordinates?

b What are G and G^{-1} in cylindrical coordinates?

7 **a** What are the lengths of the local base vectors at (2,0,0)? at (2,0,1)? at $(2,\pi/3,0)$? at $(2,\pi/3,1)$?

b What are the lengths of the local reciprocal base vectors at each of the points in part **a**?

8 If $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ are the base vectors at the point $(2,\pi/3,1)$ and if

$$\mathbf{U} = 2\mathbf{e}_1 + 3\mathbf{e}_2 + \mathbf{e}_3 \qquad \text{and} \qquad \mathbf{V} = \mathbf{e}_1 - \mathbf{e}_2 + 2\mathbf{e}_3$$

what is the length of **U**? the length of **V**? the angle between **U** and **V**?

9 Let **V** be the vector extending from the point $(2,0,1)$ to $(2,\pi/3,1)$. Express **V** in terms of the base vectors at $(2,0,1)$ and also in terms of the base vectors at $(2,\pi/3,1)$ Check the length of **V**, using each of these representations

10 Work Exercise 9 using the reciprocal base vectors at $(2,0,1)$ and at $(2,\pi/3,1)$

13.4

Tensors

In the last section, without having referred to them by name, we were already working with tensors Now, with the experience we have gained from our discussion of coordinate transformations in three dimensions, we can make the matter explicit

Let $(x^1,x^2,\ldots,x^n)$ and $(\bar{x}^1,\bar{x}^2,\ldots,\bar{x}^n)$ be generalized coordinates in n dimensions, and let the two systems of coordinates be related by the transformation equations

$$(1)\qquad \begin{array}{lll} T & \bar{x}^i = \bar{x}^i(x^1,x^2,\ldots,x^n) & \\ & & i = 1, 2, \ldots, n \\ T^{-1} & x^i = x^i(\bar{x}^1,\bar{x}^2,\ldots,\bar{x}^n) & \end{array}$$

Once we pass beyond the three-dimensional space of experience, geometric intuition is of little help to us However, it can be shown that in n dimensions, just as in three dimensions, there are n parametric curves passing through an arbitrary point and on each of these curves one and only one of the generalized coordinates varies Moreover, if the Jacobian determinant of the transformation (1) is different from zero, vectors tangent to the parametric curves through an arbitrary point can be shown to be linearly independent. Hence, if local base vectors $\mathbf{e}_i$ $(i = 1, 2, \ldots, n)$ are defined at an arbitrary point P by the conditions that

$$ds\dagger = |\mathbf{e}_i\, dx^i| = \sqrt{\mathbf{e}_i \cdot \mathbf{e}_i}\, |dx^i| \qquad i \text{ not summed}$$

any vector originating at P can be expressed as a linear combination of these vectors Furthermore, a set of independent local reciprocal base vectors $\mathbf{e}^i$ $(i = 1, 2, \ldots, n)$ can be defined at any point P by the same conditions we used in three dimensions, namely,

$$\mathbf{e}^i \cdot \mathbf{e}_j = \delta_j{}^i$$

and any vector originating at P can also be represented as a linear combination of the reciprocal base vectors at P. In fact, all the results of the last section, with the exception of those involving scalar triple products, are equally correct in n dimensions, provided only that the summation convention is understood to cover the range 1 to n instead of the range 1 to 3.

---

† Just as in three dimensions, we assume that the metrical properties of n-dimensional space are defined by a positive-definite differential quadratic form $(ds)^2 = g_{ij}\, dx^i\, dx^j$, $i,j = 1, 2, \ldots, n$, whose matrix $G = \|g_{ij}\|$ is, of course, nonsingular

By a **scalar, or tensor of rank zero,** we mean a quantity S whose descriptions in the two coordinate systems are connected by the relation

$$\bar{S}(\bar{x}^1,\bar{x}^2, \ldots ,\bar{x}^n) = S(x^1,x^2, \ldots ,x^n) \tag{2}$$

By a **contravariant vector, or contravariant tensor of rank 1,** we mean a set of n quantities ξ^i, called **components,** whose descriptions in the two coordinate systems are connected by the relations

$$\bar{\xi}^i(\bar{x}^1,\bar{x}^2, \ldots ,\bar{x}^n) = \xi^\alpha(x^1,x^2, \ldots ,x^n) \frac{\partial \bar{x}^i}{\partial x^\alpha} \qquad i = 1, 2, \ldots , n \tag{3}$$

Since $d\bar{x}^i = \dfrac{\partial \bar{x}^i}{\partial x^\alpha} dx^\alpha$, it follows that the differentials of the coordinate variables are the components of a contravariant tensor of rank 1

By a **covariant vector, or covariant tensor of rank 1,** we mean a set of n quantities ξ_i, also called **components,** whose descriptions in the two coordinate systems are connected by the relations

$$\bar{\xi}_i(\bar{x}^1,\bar{x}^2, \ldots ,\bar{x}^n) = \xi_\alpha(x^1,x^2, \ldots ,x^n) \frac{\partial x^\alpha}{\partial \bar{x}^i} \qquad i = 1, 2, \ldots , n \tag{4}$$

If ϕ is a scalar point function [for which, therefore,

$$\bar{\phi}(\bar{x}^1,\bar{x}^2, \ldots ,\bar{x}^n) = \phi(x^1,x^2, \ldots ,x^n)$$

then

$$\frac{\partial \bar{\phi}}{\partial \bar{x}^i} = \frac{\partial \phi}{\partial x^\alpha} \frac{\partial x^\alpha}{\partial \bar{x}^i}$$

Hence, the n quantities $\dfrac{\partial \phi}{\partial x^i}$ are the components of a covariant tensor of rank 1, which we recognize as the gradient of ϕ

A **contravariant tensor of rank 2** is a set of n^2 quantities ξ^{ij} whose descriptions in the two coordinate systems are connected by the relations

$$\bar{\xi}^{ij} = \xi^{\alpha\beta} \frac{\partial \bar{x}^i}{\partial x^\alpha} \frac{\partial \bar{x}^j}{\partial x^\beta} \qquad i,j = 1, 2, \ldots , n \tag{5}$$

From Eq (26), Sec 13.3, it is clear that the elements g^{ij} of the matrix G^{-1} form a contravariant tensor of rank 2.

A **covariant tensor of rank 2** is a set of n^2 quantities ξ_{ij} whose descriptions in the two coordinate systems are connected by the relations

$$\bar{\xi}_{ij} = \xi_{\alpha\beta} \frac{\partial x^\alpha}{\partial \bar{x}^i} \frac{\partial x^\beta}{\partial \bar{x}^j} \qquad i,j = 1, 2, \ldots , n \tag{6}$$

From Eq. (17), Sec. 13 3, it is clear that the elements g_{ij} of the fundamental matrix G form a covariant tensor of rank 2. This tensor is often called the **fundamental metric tensor.**

A **mixed tensor of rank 2** is a set of n^2 quantities $\xi_j^{\,i}$ whose descriptions in the two coordinate systems are connected by the relations

$$\bar\xi_j^{\,i} = \xi_\beta^{\,\alpha} \frac{\partial \bar x^i}{\partial x^\alpha} \frac{\partial x^\beta}{\partial \bar x^j} \tag{7}$$

Although we shall leave the proof of this fact as an exercise, $\delta_j^{\,i}$ is an example of a mixed tensor of rank 2

A tensor ξ^{ij} (ξ_{ij}) such that $\xi^{ij} = \xi^{ji}$ $(\xi_{ij} = \xi_{ji})$ for all values of i and j is said to be **symmetric.** A tensor ξ^{ij} (ξ_{ij}) such that $\xi^{ij} = -\xi^{ji}$ $(\xi_{ij} = -\xi_{ji})$ for all values of i and j is said to be **skew-symmetric** or **alternating.**

The concept of a tensor can, clearly, be generalized to include tensors of arbitrary rank r with any number k $(0 \leqq k \leqq r)$ of covariant indices and $r - k$ contravariant indices For instance, a set of n^5 quantities ξ^{ij}_{uvw} whose descriptions in the two coordinate systems are connected by the relations

$$\bar\xi^{ij}_{uvw} = \xi^{\alpha\beta}_{\delta\gamma\epsilon} \frac{\partial \bar x^i}{\partial x^\alpha} \frac{\partial \bar x^j}{\partial x^\beta} \frac{\partial x^\delta}{\partial \bar x^u} \frac{\partial x^\gamma}{\partial \bar x^v} \frac{\partial x^\epsilon}{\partial \bar x^w}$$

constitutes a mixed tensor of rank 5 with two contravariant indices i and j and three covariant indices u, v, and w

From the definition of a tensor as a set of quantities which transform in a prescribed way, it is clear that a tensor can be constructed by specifying its components in one coordinate system arbitrarily and then letting the appropriate transformation laws define its components in all other coordinate systems

The algebra of tensors is based primarily upon the following observations

Two tensors are equal if and only if they have the same rank and the same number of indices of each type and have their corresponding components equal in one and, hence, in all coordinate systems. In particular, *if the components of a tensor are all zero in one coordinate system, they are all zero in every coordinate system.*

If T_1 and T_2 are tensors of the same type, then the set of quantities obtained by adding respective components of T_1 and T_2 is a tensor $T_1 + T_2$ of the same type as T_1 and T_2

If T_1 is a tensor of rank r_1 with c_1 contravariant and γ_1 covariant indices and if T_2 is a tensor of rank r_2 with c_2 contravariant and γ_2 covariant indices, then the set of quantities obtained by multiplying each component of T_1 by each component of T_2 is a tensor T_1T_2, called the **outer product** of T_1 and T_2, of rank $r_1 + r_2$ with $c_1 + c_2$ contravariant indices and $\gamma_1 + \gamma_2$ covariant indices For instance, if T_1 is the tensor ξ^{ij} and T_2 is the tensor ξ_l^k, then the general term $\xi^{ij}\,\xi_l^k$ transforms according to the law

$$\bar\xi^{ij}\bar\xi_l^k = \left(\xi^{\alpha\beta}\frac{\partial\bar x^i}{\partial x^\alpha}\frac{\partial\bar x^j}{\partial x^\beta}\right)\left(\xi_\delta^\gamma\frac{\partial\bar x^k}{\partial x^\gamma}\frac{\partial x^\delta}{\partial\bar x^l}\right) = \xi^{\alpha\beta}\xi_\delta^\gamma\frac{\partial\bar x^i}{\partial x^\alpha}\frac{\partial\bar x^j}{\partial x^\beta}\frac{\partial\bar x^k}{\partial x^\gamma}\frac{\partial x^\delta}{\partial\bar x^l}$$

which shows that $T_1T_2 = \xi^{ij}\xi_l^k$ is a tensor, say η_l^{ijk}, of rank 4 with 3 contravariant and 1 covariant indices.

If, in a tensor of any type, a contravariant index is summed against a covariant index by simply setting one index equal to the other and thereby invoking the summation convention, the resulting set of quantities is a tensor with one less contravariant index and one less covariant index. For example, since the tensor ξ_k^{ij} transforms according to the law

$$\bar\xi_k^{ij} = \xi_\gamma^{\alpha\beta}\frac{\partial\bar x^i}{\partial x^\alpha}\frac{\partial\bar x^j}{\partial x^\beta}\frac{\partial x^\gamma}{\partial\bar x^k}$$

we have, setting $j = k$,

$$\begin{aligned}\bar\xi_j^{ij} &= \xi_\gamma^{\alpha\beta}\frac{\partial\bar x^i}{\partial x^\alpha}\frac{\partial\bar x^j}{\partial x^\beta}\frac{\partial x^\gamma}{\partial\bar x^j}\\ &= \xi_\gamma^{\alpha\beta}\frac{\partial\bar x^i}{\partial x^\alpha}\,\delta_\beta^\gamma \qquad \text{(by Lemma 1, Sec. 13.3)}\\ &= \xi_\gamma^{\alpha\gamma}\frac{\partial\bar x^i}{\partial x^\alpha}\end{aligned}$$

since only when $\beta = \gamma$ is $\delta_\beta^\gamma \neq 0$. Hence, ξ_j^{ij} transforms as a contravariant tensor of rank 1, that is, ξ_j^{ij} is a contravariant vector, say η^i. This process of obtaining one tensor from another is known as **contraction.** Obviously, the process of contraction can be repeated as long as there are indices of each type. When the process of contraction is applied to the product of two tensors in such a way that at each stage one of the two indices belongs to the first factor and the other to the second, the resulting tensor is said to be the **inner product** of the two tensors with respect to the given set of indices.

The converse of the last observation is also important. *A set of n^r quantities is a tensor provided an inner product of the set and an arbitrary tensor is also a tensor.* The proof of this assertion should be sufficiently clear from the argument for the special case $r = 2$. Suppose, then, that ξ_β^α is a set of n^2 quantities such that, for an arbitrary tensor of the second rank, say η_δ^γ, we have

$$\xi_\beta^\alpha\eta_\delta^\beta = \zeta_\delta^\alpha \tag{8}$$

where ζ_δ^α is a tensor. Under an arbitrary transformation of coordinates we have, of course,

$$\bar\xi_b^a\bar\eta_d^b = \bar\zeta_d^a \tag{9}$$

Now, since η_δ^β and ζ_δ^α are tensors of the indicated type, we have, by definition,

$$\bar\eta_d^b = \eta_\delta^\beta\frac{\partial\bar x^b}{\partial x^\beta}\frac{\partial x^\delta}{\partial\bar x^d} \qquad \text{and} \qquad \bar\zeta_d^a = \zeta_\delta^\alpha\frac{\partial\bar x^a}{\partial x^\alpha}\frac{\partial x^\delta}{\partial\bar x^d}$$

Hence, substituting into (9) and then using (8), we have

$$\bar{\xi}_b{}^a \eta_\delta{}^\beta \frac{\partial \bar{x}^b}{\partial x^\beta} \frac{\partial x^\delta}{\partial \bar{x}^d} = \zeta_\delta{}^\alpha \frac{\partial \bar{x}^a}{\partial x^\alpha} \frac{\partial x^\delta}{\partial \bar{x}^d} = \xi_\beta{}^\alpha \eta_\delta{}^\beta \frac{\partial \bar{x}^a}{\partial x^\alpha} \frac{\partial x^\delta}{\partial \bar{x}^d}$$

From this, by transposing and collecting terms, we obtain

$$\frac{\partial x^\delta}{\partial \bar{x}^d} \left(\bar{\xi}_b{}^a \frac{\partial \bar{x}^b}{\partial x^\beta} - \xi_\beta{}^\alpha \frac{\partial \bar{x}^a}{\partial x^\alpha} \right) \eta_\delta{}^\beta = 0$$

If we now form the inner product of the expression on the left with $\frac{\partial \bar{x}^d}{\partial x^\epsilon}$ and recall from Lemma 1, Sec 13 3, that

$$\frac{\partial x^\delta}{\partial \bar{x}^d} \frac{\partial \bar{x}^d}{\partial x^\epsilon} = \delta_\epsilon{}^\delta$$

we have
$$\delta_\epsilon{}^\delta \left(\bar{\xi}_b{}^a \frac{\partial \bar{x}^b}{\partial x^\beta} - \xi_\beta{}^\alpha \frac{\partial \bar{x}^a}{\partial x^\alpha} \right) \eta_\delta{}^\beta = 0$$

or
$$\left(\bar{\xi}_b{}^a \frac{\partial \bar{x}^b}{\partial x^\beta} - \xi_\beta{}^\alpha \frac{\partial \bar{x}^a}{\partial x^\alpha} \right) \eta_\epsilon{}^\beta = 0$$

Now, since $\eta_\epsilon{}^\beta$ is completely arbitrary, it may be chosen, in turn, to be a tensor each of whose components except one is equal to zero Hence it follows that the expression in parentheses in the last equation must be identically zero. Therefore,

$$\bar{\xi}_b{}^a \frac{\partial \bar{x}^b}{\partial x^\beta} = \xi_\beta{}^\alpha \frac{\partial \bar{x}^a}{\partial x^\alpha}$$

Finally, if we form the inner product of each member of this equation with $\frac{\partial x^\beta}{\partial \bar{x}^e}$ and again use Lemma 1, Sec 13.3, we have

$$\bar{\xi}_b{}^a \delta_e{}^b = \xi_\beta{}^\alpha \frac{\partial \bar{x}^a}{\partial x^\alpha} \frac{\partial x^\beta}{\partial \bar{x}^e}$$

or
$$\bar{\xi}_e{}^a = \xi_\beta{}^\alpha \frac{\partial \bar{x}^a}{\partial x^\alpha} \frac{\partial x^\beta}{\partial \bar{x}^e} \equiv \xi_\epsilon{}^\alpha \frac{\partial \bar{x}^a}{\partial x^\alpha} \frac{\partial x^\epsilon}{\partial \bar{x}^e}$$

which is precisely the law of transformation for a mixed tensor of rank 2. In other words, $\xi_\epsilon{}^\alpha$ is a tensor, as asserted The property we have confirmed in this particular case is often referred to as the **quotient law** for tensors

The preceding observation is frequently the most effective means of proving that a set of quantities is a tensor For instance, by its use we can establish the following interesting result *If the elements of a nonsingular matrix* $\|f_{ij}\|$ *are the components of a covariant tensor of rank* 2, *then the elements of the inverse matrix* $\|f^{ij}\|$ *are the components of a contravariant tensor of rank* 2. To prove this, let ξ^α be any contravariant vector. Then, by the process of contraction,

$$\eta_i = f_{i\alpha} \xi^\alpha$$

is a covariant vector. Moreover, since $\|f_{ij}\|$ is nonsingular and ξ^α is arbitrary, η_i is also arbitrary; that is, any covariant vector η_i can be obtained in this fashion from a suitable contravariant vector ξ^α. If we now form the inner product of each member of the last equation with $f^{\beta i}$, we have

$$f^{\beta i}\eta_i = f^{\beta i}f_{i\alpha}\xi^\alpha$$

However, since $\|f_{ij}\|$ and $\|f^{ij}\|$ are inverses, it follows that

$$f^{\beta i}f_{i\alpha} = \delta_\alpha{}^\beta$$

Hence, $\quad f^{\beta i}\eta_i = \delta_\alpha{}^\beta\xi^\alpha = \xi^\beta$

Since η_i is arbitrary, it follows from the quotient law that f^{ij} is a tensor, as asserted

EXERCISES

1 Verify that $\delta_j{}^i$ is a mixed tensor of rank 2

2 **a** Is $\delta_1{}^i$ a tensor? **b** Is $\delta_j{}^2$ a tensor?

3 Verify that, if T_1 and T_2 are tensors of the same type, then $T_1 \pm T_2$ is also a tensor of that type.

4 Verify that, if T is a tensor and ϕ is a scalar, then the set of quantities obtained by multiplying each component of T by ϕ is a tensor of the same type as T

5 **a** Is a tensor obtained if two covariant indices in a tensor are summed against each other?
b Is a tensor obtained if two contravariant indices in a tensor are summed against each other?

6 If the elements of a nonsingular matrix $\|f_j{}^i\|$ are the components of a mixed tensor of rank 2, do the elements of the inverse matrix form a tensor?

7 **a** Let $\xi_\beta{}^\alpha$ be a set of n^2 quantities, and let $\eta^{\beta\gamma}$ be an arbitrary contravariant tensor of rank 2 Show that $\xi_\beta{}^\alpha$ is a tensor if the product $\xi_\beta{}^\alpha\eta^{\beta\gamma}$ is a tensor $\zeta^{\alpha\gamma}$
b Show that $\xi_\beta{}^\alpha$ is a tensor if its inner product with an arbitrary covariant tensor is also a tensor

8 **a** Show that $\xi_{\alpha\beta}$ is a tensor if its inner product with an arbitrary mixed tensor $\eta_\gamma{}^\beta$ is a tensor
b Show that $\xi^{\alpha\beta}$ is a tensor if the product $\xi^{\alpha\beta}\eta_{\beta\gamma}$ is a tensor, $\eta_{\beta\gamma}$ being an arbitrary covariant tensor of rank 2

9 If $\eta_{\alpha\beta}^\delta$ is an arbitrary tensor of the indicated type, show that the n^3 quantities $\xi_\gamma^{\alpha\beta}$ form a tensor if the product $\xi_\gamma^{\alpha\beta}\eta_{\alpha\beta}^\delta$ is a tensor $\zeta_\gamma{}^\delta$

10 Show how the contravariant representation of a vector can be obtained from its covariant representation, and vice versa

13.5

Divergence and curl

We have already seen (Sec 13 4) that, if ϕ is a scalar point function, then $\dfrac{\partial\phi}{\partial x^i}$ is a covariant vector, which we recognized as the gradient of ϕ. We now turn our attention to the determination of the divergence and curl of vectors in generalized coordinates

For the divergence of a contravariant vector ξ^α we have the expression

$$\frac{1}{\sqrt{|G|}}\frac{\partial}{\partial x^\alpha}(\sqrt{|G|}\,\xi^\alpha) \tag{1}$$

where G is the metric tensor of the space In rectangular coordinates, for which

$$G = \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}$$

this is clearly the divergence, as defined in Sec. 12 3 However, before this can be accepted as a definition of the divergence in any coordinate system we must prove that it is a scalar invariant; that is, that it is the same in all coordinate systems To do this, let us consider the given expression in a second coordinate system·

$$\frac{1}{\sqrt{|\bar G|}}\frac{\partial}{\partial \bar x^i}\left(\sqrt{|\bar G|}\,\bar\xi^i\right) = \frac{1}{\sqrt{|\bar G|}}\left(\frac{1}{2\sqrt{|\bar G|}}\frac{\partial|\bar G|}{\partial \bar x^i}\bar\xi^i + \sqrt{|\bar G|}\frac{\partial\bar\xi^i}{\partial\bar x^i}\right)$$

$$= \frac{1}{2|\bar G|}\frac{\partial|\bar G|}{\partial x^a}\frac{\partial x^a}{\partial\bar x^i}\bar\xi^i + \frac{\partial\bar\xi^i}{\partial\bar x^i} \tag{2}$$

By hypothesis, ξ^a is a contravariant vector Hence,

$$\bar\xi^i = \xi^a\frac{\partial\bar x^i}{\partial x^a} \tag{3}$$

and

$$\bar\xi^i\frac{\partial x^a}{\partial\bar x^i} = \xi^a \tag{4}$$

Also, from Eq (19), Sec 13 3, $|\bar G| = |G|\ |J|^{-2}$, where J is the Jacobian of the transformation Therefore, by differentiating and dividing by $|\bar G|$, we obtain

$$\frac{1}{|\bar G|}\frac{\partial|\bar G|}{\partial x^a} = \frac{1}{|G|}\frac{\partial|G|}{\partial x^a} - \frac{2}{|J|}\frac{\partial|J|}{\partial x^a} \tag{5}$$

Thus, substituting from Eqs (3), (4), and (5) into Eq (2), we have

$$\frac{1}{\sqrt{|\bar G|}}\frac{\partial}{\partial\bar x^i}\left(\sqrt{|\bar G|}\,\bar\xi^i\right) = \frac{1}{2}\left(\frac{1}{|G|}\frac{\partial|G|}{\partial x^a} - \frac{2}{|J|}\frac{\partial|J|}{\partial x^a}\right)\xi^a + \frac{\partial}{\partial\bar x^i}\left(\xi^a\frac{\partial\bar x^i}{\partial x^a}\right)$$

$$= \frac{1}{2|G|}\frac{\partial|G|}{\partial x^a}\xi^a - \frac{1}{|J|}\frac{\partial|J|}{\partial x^a}\xi^a + \frac{\partial\xi^a}{\partial\bar x^i}\frac{\partial\bar x^i}{\partial x^a} + \xi^a\frac{\partial^2\bar x^i}{\partial x^b\,\partial x^a}\frac{\partial x^b}{\partial\bar x^i}$$

$$= \left(\frac{1}{2|G|}\frac{\partial|G|}{\partial x^a}\xi^a + \frac{\partial\xi^a}{\partial x^a}\right) + \left(\frac{\partial^2\bar x^i}{\partial x^a\,\partial x^b}\frac{\partial x^b}{\partial\bar x^i} - \frac{1}{|J|}\frac{\partial|J|}{\partial x^a}\right)\xi^a \tag{6}$$

Now, the first quantity in parentheses in the last expression is precisely

$$\frac{1}{\sqrt{|G|}}\frac{\partial}{\partial x^a}\left(\sqrt{|G|}\,\xi^a\right)$$

Hence, our proof will be complete if we can show that the second quantity in parentheses is zero. To do this, we recall from the rule for differentiating determinants that the derivative of $|J|$ is equal to the sum of n determinants, the ith one of which is identical with $|J|$ except that the ith row consists of the derivatives of the elements in the ith row of $|J|$. Hence, if we denote by

J_i^b the cofactor of the element in the ith row and bth column of $|J|$, then, in expanded form,

$$\frac{\partial |J|}{\partial x^a} = \frac{\partial^2 \bar{x}^i}{\partial x^a\, \partial x^b} J_i^b$$

However, by Corollary 1, Theorem 1, Sec. 13.3,

$$J_i^b = \frac{\partial x^b}{\partial \bar{x}^i} |J|$$

Hence, $$\frac{1}{|J|}\frac{\partial |J|}{\partial x^a} = \frac{\partial^2 \bar{x}^i}{\partial x^a\, \partial x^b}\frac{\partial x^b}{\partial \bar{x}^i}$$

and, thus, the second quantity in parentheses in (6) is indeed zero; and the scalar

$$\frac{1}{\sqrt{|G|}}\frac{\partial}{\partial x^a}(\sqrt{|G|}\,\xi^a)$$

is invariant and, hence, equal to the divergence in every coordinate system.

If we use the covariant representation ξ_a instead of the contravariant representation ξ^a, then, since

$$\xi^a = g^{ab}\xi_b$$

(see Exercise 10, Sec. 13.4), we have for the divergence

$$\frac{1}{\sqrt{|G|}}\frac{\partial}{\partial x^a}(\sqrt{|G|}\,g^{ab}\xi_b) \tag{7}$$

If ξ_a is the gradient of a scalar point function, that is, if ξ_a is the covariant vector $\frac{\partial \phi}{\partial x^b}$, then its divergence is called the **Laplacian** of ϕ. In other words, in generalized coordinates,

$$\nabla^2\phi = \frac{1}{\sqrt{|G|}}\frac{\partial}{\partial x^a}\left(\sqrt{|G|}\,g^{ab}\frac{\partial \phi}{\partial x^b}\right) \tag{8}$$

EXAMPLE 1

Obtain the expression for $\nabla^2\phi$ in cylindrical coordinates

By direct calculation, as in Example 1, Sec. 13 3, or by observing from a figure that

$$(ds)^2 = (dr)^2 + (r\,d\theta)^2 + (dz)^2$$

we find that, in cylindrical coordinates,

$$G = \|g_{ij}\| = \begin{Vmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & 1 \end{Vmatrix} \qquad \text{and} \qquad G^{-1} = \|g^{ij}\| = \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1/r^2 & 0 \\ 0 & 0 & 1 \end{Vmatrix}$$

Therefore, from (8),

$$\begin{aligned}\nabla^2\phi &= \frac{1}{r}\left[\frac{\partial}{\partial r}\left(r\frac{\partial\phi}{\partial r}\right) + \frac{\partial}{\partial\theta}\left(r\frac{1}{r^2}\frac{\partial\phi}{\partial\theta}\right) + \frac{\partial}{\partial z}\left(r\frac{\partial\phi}{\partial z}\right)\right] \\ &= \frac{1}{r}\left[\left(r\frac{\partial^2\phi}{\partial r^2} + \frac{\partial\phi}{\partial r}\right) + \frac{1}{r}\frac{\partial^2\phi}{\partial\theta^2} + r\frac{\partial^2\phi}{\partial z^2}\right] \\ &= \frac{\partial^2\phi}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2\phi}{\partial\theta^2} + \frac{\partial^2\phi}{\partial z^2} + \frac{1}{r}\frac{\partial\phi}{\partial r}\end{aligned}$$

Consider, now, an arbitrary covariant vector ξ_a. From its law of transformation

$$\xi_\alpha = \xi_a \frac{\partial x^a}{\partial \bar{x}^\alpha}$$

we have, by differentiation,

$$\frac{\partial \xi_\alpha}{\partial \bar{x}^\beta} = \frac{\partial \xi_a}{\partial x^b}\frac{\partial x^b}{\partial \bar{x}^\beta}\frac{\partial x^a}{\partial \bar{x}^\alpha} + \xi_a \frac{\partial^2 x^a}{\partial \bar{x}^\beta\, \partial \bar{x}^\alpha}$$

Similarly, of course,

$$\frac{\partial \xi_\beta}{\partial \bar{x}^\alpha} = \frac{\partial \xi_b}{\partial x^a}\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\beta} + \xi_b \frac{\partial^2 x^b}{\partial \bar{x}^\alpha\, \partial \bar{x}^\beta}$$

Hence, subtracting the last two equations, we have

$$\frac{\partial \xi_\alpha}{\partial \bar{x}^\beta} - \frac{\partial \xi_\beta}{\partial \bar{x}^\alpha} = \left(\frac{\partial \xi_a}{\partial x^b} - \frac{\partial \xi_b}{\partial x^a}\right)\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\beta}$$

where the other terms cancel, since the order of partial differentiation is immaterial and since a and b are just dummy indices From the law of transformation defined by the last equation, it follows that

$$\frac{\partial \xi_a}{\partial x^b} - \frac{\partial \xi_b}{\partial x^a}$$

is a covariant tensor of the second rank. Clearly, it is a generalization of the familiar notion of the curl of a vector.

More specifically, in three dimensions, let ξ_{ab} be an arbitrary alternating covariant tensor of the second rank, for which, by definition, $\xi_{ab} = -\xi_{ba}$. From ξ_{ab} we can construct the expression

$$\tfrac{1}{2}\xi_{ab}\mathbf{e}^a \times \mathbf{e}^b \equiv \xi_{12}\mathbf{e}^1 \times \mathbf{e}^2 + \xi_{23}\mathbf{e}^2 \times \mathbf{e}^3 + \xi_{31}\mathbf{e}^3 \times \mathbf{e}^1$$

Moreover, if we use the fact (see Exercise 1, below) that

$$\mathbf{e}^i \times \mathbf{e}^j = \frac{\mathbf{e}_k}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]} \qquad i,j,k \text{ any cyclic permutation of } 1,2,3$$

we can write $\frac{1}{2}\xi_{ab}\mathbf{e}^a \times \mathbf{e}^b$ in the form

$$\xi_{12}\frac{\mathbf{e}_3}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]} + \xi_{23}\frac{\mathbf{e}_1}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]} + \xi_{31}\frac{\mathbf{e}_2}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]}$$

from which it is clear that $\frac{1}{2}\xi_{ab}\mathbf{e}^a \times \mathbf{e}^b$ is a contravariant tensor of rank 1 Finally, if we recall from Eq (12), Sec. 13 3, that

$$[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]^2 = |G|$$

we can write this tensor in the form

(9)
$$\frac{\xi_{23}}{-\sqrt{|G|}}\mathbf{e}_1 + \frac{\xi_{31}}{-\sqrt{|G|}}\mathbf{e}_2 + \frac{\xi_{12}}{-\sqrt{|G|}}\mathbf{e}_3$$

If $\xi_{ab} = \dfrac{\partial \xi_a}{\partial x^b} - \dfrac{\partial \xi_b}{\partial x^a}$, where ξ_a is a covariant vector, then the expression (9), with the negative square root used, as indicated, is precisely the curl of ξ_a, as we defined it in Chap. 12, Sec. 12.3.

EXERCISES

1 Using Eqs. (9) and (10), Sec. 13.3, or otherwise, show that

$$\mathbf{e}^2 \times \mathbf{e}^3 = \frac{\mathbf{e}_1}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]} \qquad \mathbf{e}^3 \times \mathbf{e}^1 = \frac{\mathbf{e}_2}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]} \qquad \mathbf{e}^1 \times \mathbf{e}^2 = \frac{\mathbf{e}_3}{[\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3]}$$

2 a What is the divergence of a contravariant vector in cylindrical coordinates?
b What is the divergence of a covariant vector in cylindrical coordinates?

3 a What is the divergence of a contravariant vector in spherical coordinates?
b What is the divergence of a covariant vector in spherical coordinates?

4 Obtain the expression for $\nabla^2\phi$ in spherical coordinates.

5 If ξ_a is the gradient of a scalar function ϕ, show that the curl of ξ_a vanishes identically

13.6

Covariant differentiation

Since the components of a tensor are functions of the generalized coordinates, it is obvious that they can be differentiated partially with respect to the coordinate variables However, the quantities thus obtained are of no intrinsic interest since they are not the components of a tensor. For instance, if ξ^d is a contravariant vector and if we differentiate the transformation equation

$$\bar{\xi}^\delta = \xi^d \frac{\partial \bar{x}^\delta}{\partial x^d}$$

partially with respect to $\bar{x}^\beta$, we have

$$\frac{\partial \bar{\xi}^\delta}{\partial \bar{x}^\beta} = \frac{\partial \xi^d}{\partial x^b}\frac{\partial x^b}{\partial \bar{x}^\beta}\frac{\partial \bar{x}^\delta}{\partial x^d} + \xi^d \frac{\partial^2 \bar{x}^\delta}{\partial x^b\,\partial x^d}\frac{\partial x^b}{\partial \bar{x}^\beta} \tag{1}$$

Clearly, if the second term on the right were not present, $\dfrac{\partial \xi^d}{\partial x^b}$ would be a mixed tensor of rank 2, since the first term on the right in (1) is precisely what is given by the law of transformation for a tensor with one covariant and one contravariant index. It is also interesting to note that, if the equations connecting the two sets of generalized coordinates are linear, as they are for transformations between rectangular and oblique coordinates, then the second term *is* missing. These observations raise the important question of whether or not it is possible to add "correction" terms $C_b{}^d$ to the partial derivatives $\dfrac{\partial \xi^d}{\partial x^b}$ so that

$$\frac{\partial \xi^d}{\partial x^b} + C_b{}^d$$

will be a mixed tensor of rank 2. This is indeed possible, and, although we cannot go deeply into the matter, we shall determine the appropriate "correction" terms and define the so-called *covariant derivative*.

From Eq (1) it is almost obvious that the terms to be added to $\dfrac{\partial \xi^d}{\partial x^b}$ to eliminate the second sum should be linear in the ξ's, say,

$$C_b{}^d = \Gamma_{ab}{}^d \xi^a$$

and this is actually the case. To determine the coefficient function $\Gamma_{ab}{}^d$, we begin with the metric tensor g_{ab} From its law of transformation, namely,

$$\bar{g}_{\alpha\beta} = g_{ab}\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\beta}$$

by differentiating each side with respect to $\bar{x}^\gamma$, we obtain

(2) $$\frac{\partial \bar{g}_{\alpha\beta}}{\partial \bar{x}^\gamma} = \frac{\partial g_{ab}}{\partial x^c}\frac{\partial x^c}{\partial \bar{x}^\gamma}\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\beta} + \left[g_{ab}\frac{\partial^2 x^a}{\partial \bar{x}^\gamma\,\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\beta}\right] + \left\{g_{ab}\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial^2 x^b}{\partial \bar{x}^\gamma\,\partial \bar{x}^\beta}\right\}$$

From this, by first interchanging β and γ and then interchanging γ and α, and making the corresponding permutations of the dummy indices a, b, c in the first term, we obtain, respectively,

(3) $$\frac{\partial \bar{g}_{\alpha\gamma}}{\partial \bar{x}^\beta} = \frac{\partial g_{ac}}{\partial x^b}\frac{\partial x^b}{\partial \bar{x}^\beta}\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^c}{\partial \bar{x}^\gamma} + g_{ab}\frac{\partial^2 x^a}{\partial \bar{x}^\beta\,\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\gamma} + \left\{g_{ab}\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial^2 x^b}{\partial \bar{x}^\beta\,\partial \bar{x}^\gamma}\right\}$$

(4) $$\frac{\partial \bar{g}_{\gamma\beta}}{\partial \bar{x}^\alpha} = \frac{\partial g_{cb}}{\partial x^a}\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^c}{\partial \bar{x}^\gamma}\frac{\partial x^b}{\partial \bar{x}^\beta} + \left[g_{ab}\frac{\partial^2 x^a}{\partial \bar{x}^\alpha\,\partial \bar{x}^\gamma}\frac{\partial x^b}{\partial \bar{x}^\beta}\right] + g_{ab}\frac{\partial x^a}{\partial \bar{x}^\gamma}\frac{\partial^2 x^b}{\partial \bar{x}^\alpha\,\partial \bar{x}^\beta}$$

Now, subtracting (2) from the sum of (3) and (4), noting that the quantities in brackets and the quantities in braces cancel respectively, we obtain

$$\frac{\partial \bar{g}_{\gamma\beta}}{\partial \bar{x}^\alpha} + \frac{\partial \bar{g}_{\alpha\gamma}}{\partial \bar{x}^\beta} - \frac{\partial \bar{g}_{\alpha\beta}}{\partial \bar{x}^\gamma} = \left(\frac{\partial g_{cb}}{\partial x^a} + \frac{\partial g_{ac}}{\partial x^b} - \frac{\partial g_{ab}}{\partial x^c}\right)\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\beta}\frac{\partial x^c}{\partial \bar{x}^\gamma}$$
$$+ g_{ab}\frac{\partial^2 x^a}{\partial \bar{x}^\beta\,\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\gamma} + g_{ab}\frac{\partial x^a}{\partial \bar{x}^\gamma}\frac{\partial^2 x^b}{\partial \bar{x}^\alpha\,\partial \bar{x}^\beta}$$

Finally, interchanging the dummy indices a and b in the last term and recalling that $g_{ba} = g_{ab}$, we have

(5) $$\frac{\partial \bar{g}_{\gamma\beta}}{\partial \bar{x}^\alpha} + \frac{\partial \bar{g}_{\alpha\gamma}}{\partial \bar{x}^\beta} - \frac{\partial \bar{g}_{\alpha\beta}}{\partial \bar{x}^\gamma} = \left(\frac{\partial g_{cb}}{\partial x^a} + \frac{\partial g_{ac}}{\partial x^b} - \frac{\partial g_{ab}}{\partial x^c}\right)\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\beta}\frac{\partial x^c}{\partial \bar{x}^\gamma} + 2g_{ab}\frac{\partial^2 x^a}{\partial \bar{x}^\alpha\,\partial \bar{x}^\beta}\frac{\partial x^b}{\partial \bar{x}^\gamma}$$

The quantities

(6) $$\Gamma_{c,ab} = \frac{1}{2}\left(\frac{\partial g_{cb}}{\partial x^a} + \frac{\partial g_{ac}}{\partial x^b} - \frac{\partial g_{ab}}{\partial x^c}\right)$$

whose law of transformation is given by Eq. (5), are known as **Christoffel symbols of the first kind.*** Incidentally, because of the second term on the right in the transformation equation (5), it is clear that $\Gamma_{c,ab}$ is *not* a tensor.

The **Christoffel symbols of the second kind** are, by definition, the quantities

(7) $$\Gamma_{ab}{}^d = g^{dc}\Gamma_{c,ab}$$

---

* Named for the German mathematician E. B. Christoffel (1829–1900).

To obtain their law of transformation, we begin by recalling from Eq. (26), Sec. 13.3, that

$$\bar{g}^{\delta\gamma} = g^{d\iota}\frac{\partial \bar{x}^\delta}{\partial x^d}\frac{\partial \bar{x}^\gamma}{\partial x^\iota}$$

Hence,

$$\begin{aligned}\Gamma_{\alpha\beta}{}^{\delta} &= \bar{g}^{\delta\gamma}\Gamma_{\gamma,\alpha\beta} = \frac{1}{2}\bar{g}^{\delta\gamma}\left(\frac{\partial \bar{g}_{\gamma\beta}}{\partial \bar{x}^\alpha} + \frac{\partial \bar{g}_{\alpha\gamma}}{\partial \bar{x}^\beta} - \frac{\partial \bar{g}_{\alpha\beta}}{\partial \bar{x}^\gamma}\right)\\
&= \frac{1}{2}\left(g^{d\iota}\frac{\partial \bar{x}^\delta}{\partial x^d}\frac{\partial \bar{x}^\gamma}{\partial x^\iota}\right)\left[\left(\frac{\partial g_{cb}}{\partial x^a} + \frac{\partial g_{ac}}{\partial x^b} - \frac{\partial g_{ab}}{\partial x^c}\right)\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\beta}\frac{\partial x^c}{\partial \bar{x}^\gamma}\right.\\
&\qquad\left. + 2g_{ab}\frac{\partial^2 x^a}{\partial \bar{x}^\alpha\,\partial \bar{x}^\beta}\frac{\partial x^b}{\partial \bar{x}^\gamma}\right]\\
&= \frac{1}{2}g^{d\iota}\left(\frac{\partial g_{cb}}{\partial x^a} + \frac{\partial g_{ac}}{\partial x^b} - \frac{\partial g_{ab}}{\partial x^c}\right)\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\beta}\frac{\partial \bar{x}^\delta}{\partial x^d}\left[\frac{\partial x^c}{\partial \bar{x}^\gamma}\frac{\partial \bar{x}^\gamma}{\partial x^\iota}\right]\\
&\qquad + g^{d\iota}g_{ab}\frac{\partial^2 x^a}{\partial \bar{x}^\alpha\,\partial \bar{x}^\beta}\frac{\partial \bar{x}^\delta}{\partial x^d}\left[\frac{\partial x^b}{\partial \bar{x}^\gamma}\frac{\partial \bar{x}^\gamma}{\partial x^\iota}\right]\end{aligned}$$

Now, by Lemma 1, Sec. 13.3, the bracketed terms become

$$\frac{\partial x^c}{\partial \bar{x}^\gamma}\frac{\partial \bar{x}^\gamma}{\partial x^\iota} = \delta_\iota{}^c \qquad \text{and} \qquad \frac{\partial x^b}{\partial \bar{x}^\gamma}\frac{\partial \bar{x}^\gamma}{\partial x^\iota} = \delta_\iota{}^b$$

Therefore, the last equation simplifies to

$$\Gamma_{\alpha\beta}{}^{\delta} = \frac{1}{2}g^{dc}\left(\frac{\partial g_{cb}}{\partial x^a} + \frac{\partial g_{ac}}{\partial x^b} - \frac{\partial g_{ab}}{\partial x^c}\right)\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\beta}\frac{\partial \bar{x}^\delta}{\partial x^d} + g^{db}g_{ab}\frac{\partial^2 x^a}{\partial \bar{x}^\alpha\,\partial \bar{x}^\beta}\frac{\partial \bar{x}^\delta}{\partial x^d}$$

Furthermore, since $\|g^{ij}\|$ and $\|g_{ij}\|$ are inverse matrices, it follows that

$$g^{db}g_{ab} = g^{db}g_{ba} = \delta_a{}^d$$

Hence, the last term in the preceding equation reduces to

$$\frac{\partial^2 x^a}{\partial \bar{x}^\alpha\,\partial \bar{x}^\beta}\frac{\partial x^\delta}{\partial x^a}$$

and we have, finally, the law of transformation

$$\begin{aligned}\Gamma_{\alpha\beta}{}^{\delta} &= \frac{1}{2}g^{dc}\left(\frac{\partial g_{cb}}{\partial x^a} + \frac{\partial g_{ac}}{\partial x^b} - \frac{\partial g_{ab}}{\partial x^c}\right)\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\beta}\frac{\partial \bar{x}^\delta}{\partial x^d} + \frac{\partial^2 x^a}{\partial \bar{x}^\alpha\,\partial \bar{x}^\beta}\frac{\partial \bar{x}^\delta}{\partial x^a}\\
&= g^{dc}\Gamma_{c,ab}\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\beta}\frac{\partial \bar{x}^\delta}{\partial x^d} + \frac{\partial^2 x^a}{\partial \bar{x}^\alpha\,\partial \bar{x}^\beta}\frac{\partial \bar{x}^\delta}{\partial x^a}\\
&= \Gamma_{ab}{}^{d}\frac{\partial x^a}{\partial \bar{x}^\alpha}\frac{\partial x^b}{\partial \bar{x}^\beta}\frac{\partial \bar{x}^\delta}{\partial x^d} + \frac{\partial^2 x^a}{\partial \bar{x}^\alpha\,\partial \bar{x}^\beta}\frac{\partial \bar{x}^\delta}{\partial x^a} \qquad (8)\end{aligned}$$

Because of the second term on the right in (8), it is clear that $\Gamma_{ab}{}^d$, like $\Gamma_{c,ab}$, is *not* a tensor.

We can now establish the fundamental result that $\dfrac{\partial \xi^d}{\partial x^b} + \Gamma_{ab}{}^d\xi^a$ is a mixed tensor of rank 2. In fact, knowing the law of transformation for $\dfrac{\partial \xi^d}{\partial x^b}$, namely, Eq. (1), and the law of transformation for

$\Gamma_{ab}{}^{d}$, namely, Eq. (8), we have

$$\frac{\partial \xi^{\delta}}{\partial \bar{x}^{\beta}} + \Gamma_{\alpha\beta}{}^{\delta}\xi^{\alpha} = \frac{\partial \xi^{d}}{\partial x^{b}}\frac{\partial x^{b}}{\partial \bar{x}^{\beta}}\frac{\partial \bar{x}^{\delta}}{\partial x^{d}} + \xi^{d}\frac{\partial^{2}\bar{x}^{\delta}}{\partial x^{b}\,\partial x^{d}}\frac{\partial x^{b}}{\partial \bar{x}^{\beta}}$$
$$+ \left(\Gamma_{ab}{}^{d}\frac{\partial x^{a}}{\partial \bar{x}^{\alpha}}\frac{\partial x^{b}}{\partial \bar{x}^{\beta}}\frac{\partial \bar{x}^{\delta}}{\partial x^{d}} + \frac{\partial^{2}x^{a}}{\partial \bar{x}^{\alpha}\,\partial \bar{x}^{\beta}}\frac{\partial \bar{x}^{\delta}}{\partial x^{a}}\right)\xi^{i}\frac{\partial \bar{x}^{\alpha}}{\partial x^{i}}$$

or, replacing the dummy index d by i in the second term, replacing the dummy index a by b in the fourth term, and observing that in the third term $\frac{\partial x^{a}}{\partial \bar{x}^{\alpha}}\frac{\partial \bar{x}^{\alpha}}{\partial x^{i}} = \delta_{i}{}^{a}$,

$$\frac{\partial \xi^{\delta}}{\partial \bar{x}^{\beta}} + \Gamma_{\alpha\beta}{}^{\delta}\xi^{\alpha} = \frac{\partial \xi^{d}}{\partial x^{b}}\frac{\partial x^{b}}{\partial \bar{x}^{\beta}}\frac{\partial \bar{x}^{\delta}}{\partial x^{d}} + \xi^{i}\frac{\partial^{2}\bar{x}^{\delta}}{\partial x^{b}\,\partial x^{i}}\frac{\partial x^{b}}{\partial \bar{x}^{\beta}}$$
$$+ \xi^{i}\Gamma_{ab}{}^{d}\,\delta_{i}{}^{a}\frac{\partial x^{b}}{\partial \bar{x}^{\beta}}\frac{\partial \bar{x}^{\delta}}{\partial x^{d}} + \xi^{i}\frac{\partial^{2}x^{b}}{\partial \bar{x}^{\alpha}\,\partial \bar{x}^{\beta}}\frac{\partial \bar{x}^{\delta}}{\partial x^{b}}\frac{\partial \bar{x}^{\alpha}}{\partial x^{i}}$$
$$(9) \qquad = \left(\frac{\partial \xi^{d}}{\partial x^{b}} + \xi^{a}\Gamma_{ab}{}^{d}\right)\frac{\partial x^{b}}{\partial \bar{x}^{\beta}}\frac{\partial \bar{x}^{\delta}}{\partial x^{d}}$$
$$+ \xi^{i}\left[\frac{\partial^{2}\bar{x}^{\delta}}{\partial x^{b}\,\partial x^{i}}\frac{\partial x^{b}}{\partial \bar{x}^{\beta}} + \frac{\partial^{2}x^{b}}{\partial \bar{x}^{\alpha}\,\partial \bar{x}^{\beta}}\frac{\partial \bar{x}^{\delta}}{\partial x^{b}}\frac{\partial \bar{x}^{\alpha}}{\partial x^{i}}\right]$$

Now, $\frac{\partial \bar{x}^{\delta}}{\partial x^{b}}\frac{\partial x^{b}}{\partial \bar{x}^{\beta}} = \delta_{\beta}{}^{\delta}$. Hence, differentiating with respect to x^{i}, we have

$$\frac{\partial^{2}\bar{x}^{\delta}}{\partial x^{i}\,\partial x^{b}}\frac{\partial x^{b}}{\partial \bar{x}^{\beta}} + \frac{\partial \bar{x}^{\delta}}{\partial x^{b}}\frac{\partial^{2}x^{b}}{\partial \bar{x}^{\alpha}\,\partial \bar{x}^{\beta}}\frac{\partial \bar{x}^{\alpha}}{\partial x^{i}} = 0$$

Therefore, the expression in brackets in Eq (9) is equal to zero, and we have

$$\frac{\partial \xi^{\delta}}{\partial \bar{x}^{\beta}} + \Gamma_{\alpha\beta}{}^{\delta}\xi^{\alpha} = \left(\frac{\partial \xi^{d}}{\partial x^{b}} + \Gamma_{ab}{}^{d}\xi^{a}\right)\frac{\partial x^{b}}{\partial \bar{x}^{\beta}}\frac{\partial \bar{x}^{\delta}}{\partial x^{d}}$$

which proves that

$$(10) \qquad \frac{\partial \xi^{d}}{\partial x^{b}} + \Gamma_{ab}{}^{d}\xi^{a}$$

is a mixed tensor of the second rank, as asserted

The expression (10) is called the **covariant derivative of the contravariant vector** ξ^{d}, and is frequently denoted by the symbol

$$\frac{D\xi^{d}}{\partial x^{b}}$$

In very much the same way it can be shown that, if ξ_{d} is a covariant vector, then

$$(11) \qquad \frac{\partial \xi_{d}}{\partial x^{b}} - \Gamma_{db}{}^{a}\xi_{a}$$

is a mixed tensor of rank 2. This expression is known as the **covariant derivative of the covariant vector** ξ_{d}, and is denoted by the symbol

$$\frac{D\xi_{d}}{\partial x^{b}}$$

It can also be shown that any tensor has a covariant derivative, in which a term like the second term in (10) enters for each contravariant index in the tensor and a term like the second term in (11) for each covariant index Thus, for tensors of the second rank, we have the formulas

$$\frac{D\xi^{de}}{\partial x^b} = \frac{\partial \xi^{de}}{\partial x^b} + \Gamma_{ib}{}^d\xi^{ie} + \Gamma_{ib}{}^e\xi^{di} \tag{12}$$

$$\frac{D\xi_e{}^d}{\partial x^b} = \frac{\partial \xi_e{}^d}{\partial x^b} + \Gamma_{ib}{}^d\xi_e{}^i - \Gamma_{eb}{}^i\xi_i{}^d \tag{13}$$

$$\frac{D\xi_{de}}{\partial x^b} = \frac{\partial \xi_{de}}{\partial x^b} - \Gamma_{db}{}^i\xi_{ie} - \Gamma_{eb}{}^i\xi_{di} \tag{14}$$

EXERCISES

1 a Show that $\Gamma_{d,ab} = \Gamma_{d,ba}$ b Show that $\Gamma_{ab}{}^d = \Gamma_{ba}{}^d$.

c Show that $\Gamma_{d,ab} + \Gamma_{a,bd} = \dfrac{\partial g_{ad}}{\partial x^b}$.

d Show that a necessary and sufficient condition that the Christoffel symbols all be zero is that the g_{ij}'s be constants

2 a Calculate the Christoffel symbols for a cylindrical coordinate system

b Calculate the Christoffel symbols for a spherical coordinate system

3 If ϕ is a scalar function and ξ^d is a contravariant vector, show that

$$\frac{D(\phi\xi^d)}{\partial x^b} = \frac{\partial\phi}{\partial x^b}\xi^d + \phi\frac{D\xi^d}{\partial x^b}$$

4 Prove that $\dfrac{Dg_{ij}}{\partial x^k} = 0$.

5 Prove that $\dfrac{D(\xi^a\eta_b)}{\partial x^c} = \dfrac{D\xi^a}{\partial x^c}\eta_b + \xi^a\dfrac{D\eta_b}{\partial x^c}$

CHAPTER FOURTEEN

Analytic Functions of a Complex Variable

14.1

Introduction

In our work up to this point we have frequently found the use of complex numbers necessary or at least convenient. For instance, we encountered them in the solution of linear differential equations with constant coefficients in Chap 2. In Chap. 5 they appeared in the complex impedance, which we found of considerable utility in the determination of the steady-state behavior of electrical circuits. Then, in Chap 6, their use led to the important complex exponential form of Fourier series and ultimately to the inversion integral of Laplace transform theory. Finally, in Chap. 9, we found that certain important physical problems required the consideration of Bessel functions of complex arguments

None of these applications, with the exception of the inversion integral, for which fortunately we had no immediate need, required any knowledge of the properties of complex numbers or of functions of a complex variable beyond what is ordinarily acquired in courses in college algebra and calculus There are, however, large areas of applied mathematics in which familiarity with the theory of functions of a complex variable beyond this minimum is indispensable In this and the next three chapters we shall develop the major features of this theory and illustrate some of its more striking applications

14.2

Algebraic preliminaries

By a **complex number** we mean a number of the form

$$z = x + iy$$

where x and y are real numbers and i is the so-called **imaginary** unit whose existence is postulated such that $i^2 = -1$. The real number x is called the **real component** or **real part** of z The real number y is called the **imaginary component** or **imaginary part** of z. The real and imaginary parts of a complex number or expression z are often denoted by the respective symbols

$$\mathcal{R}(z) \qquad \text{and} \qquad \mathcal{I}(z)$$

It is important to keep in mind that $\mathcal{I}(z)$, as here defined, is a real quantity.

Two complex numbers $a + ib$ and $c + id$ are said to be **equal** if and only if the real and imaginary parts of the first are, respectively, equal to the real and imaginary parts of the second. In particular, the vanishing of a complex number implies not one but two conditions, namely, that both the real part and the imaginary part of the given number are zero.

EXAMPLE 1

If $(x + y + 2) + (x^2 + y)i = 0$

then $x + y + 2 = 0$ and also $x^2 + y = 0$

From this pair of simultaneous equations it follows necessarily that

$$x = 2 \quad \text{and} \quad y = -4 \qquad \text{or} \qquad x = -1 \quad \text{and} \quad y = -1$$

If $z = x + iy$, then the **negative** of z is the complex number

$$-z = -x - iy$$

If two complex numbers differ only in the sign of their imaginary parts, either one is said to be the **conjugate** of the other The conjugate of a complex number z is usually written $\bar{z}$ or, less frequently, z^*.

Addition, subtraction, and multiplication of complex numbers follow the familiar rules for real quantities, with the additional provision that in multiplication all powers of i are to be reduced as far as possible by applying the definitive property of i and its obvious extensions.

$$i^2 = -1$$

$$i^3 = i^2 i = -i$$

$$i^4 = i^2 i^2 = 1$$

$$i^5 = i^4 i = i$$

$$\cdots\cdots\cdots$$

Thus $\quad (a + ib) \pm (c + id) = (a \pm c) + (b \pm d)i$

and $\quad (a + ib)(c + id) = (ac - bd) + (bc + ad)i$

Division of complex numbers is defined as the inverse of multiplication; that is, $(a + ib)/(c + id)$ is the complex number $z = x + iy$ which satisfies the equation $(c + id)(x + iy) = a + ib$. Performing the indicated multiplication, we find

$$(cx - dy) + (dx + cy)i = a + ib$$

which implies that

$$cx - dy = a \qquad \text{and} \qquad dx + cy = b$$

Solving these for x and y, we obtain

$$x = \frac{ac + bd}{c^2 + d^2} \qquad \text{and} \qquad y = \frac{bc - ad}{c^2 + d^2}$$

Hence, $$\frac{a + ib}{c + id} = \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}i$$

In practice, the quotient of two complex numbers is usually found by multiplying both numerator and denominator by the conjugate of the denominator

$$\frac{a + ib}{c + id} = \frac{a + ib}{c + id} \cdot \frac{c - id}{c - id} = \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}i \qquad c + id \neq 0$$

Conjugate complex numbers have various simple though important properties. For instance, if $z = x + iy$, then

(1) $$z\bar{z} = (x + iy)(x - iy) = x^2 + y^2$$

which is a purely real quantity. This is the basis for the use of conjugates in division. Also,

$$z + \bar{z} = (x + iy) + (x - iy) = 2x = 2\mathcal{R}(z)$$

or

(2) $$\mathcal{R}(z) = \frac{z + \bar{z}}{2}$$

and $$z - \bar{z} = (x + iy) - (x - iy) = 2iy = 2i\mathcal{I}(z)$$

or

(3) $$\mathcal{I}(z) = \frac{z - \bar{z}}{2i}$$

In taking the conjugate of a complicated expression, the following results are of great utility.

(4) $$\overline{z_1 \pm z_2} = \bar{z}_1 \pm \bar{z}_2$$

(5) $$\overline{z_1 z_2} = \bar{z}_1\bar{z}_2$$

(6) $$\overline{\left(\frac{z_1}{z_2}\right)} = \frac{\bar{z}_1}{\bar{z}_2} \qquad z_2 \neq 0$$

The proofs of these all follow immediately from the four laws of operation and the definition of conjugates.

EXERCISES

1 Prove that, if a number is equal to its conjugate, it is necessarily real.

2 Prove that any number is equal to the conjugate of its conjugate

3 Prove that, if the product of two complex numbers is zero, at least one of the numbers must be zero.

the other and completing the triangle thus formed, a third complex number is determined whose components are precisely those of the sum $z_1 + z_2$. Figure 14.3*b* shows the construction for the difference of two complex numbers, i.e., for the sum $z_1 + (-z_2)$ Evidently, $z_1 - z_2$ is identical in length and direction with the vector drawn from the end of z_2 to the end of z_1.

Both the sum and the difference of two complex numbers can be described in terms of the parallelogram having the given numbers for adjacent sides, for the sum is simply the diagonal of the parallelogram which passes through the common origin of the two vectors, and the difference is just the other diagonal, properly directed. Much of the utility of complex numbers in elementary engineering applications stems from the fact that they add according to the parallelogram law Since this is the experimentally established law for the addition of such things as forces, velocities, currents, and voltages, it is evident that in two dimensions complex numbers, like ordinary vectors in three dimensions, can conveniently be used to represent such quantities

Although we shall have no occasion to use it, a graphical process for multiplying and dividing complex numbers can also be devised It is based upon the following exceedingly important considerations If we have two complex numbers given in polar form, their product can be written

$$\begin{aligned} z_1z_2 &= [r_1(\cos\theta_1 + i\sin\theta_1)][r_2(\cos\theta_2 + i\sin\theta_2)] \\ &= r_1r_2[(\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + i(\sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2)] \\ &= r_1r_2[\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)] \qquad (4) \end{aligned}$$

and their quotient can be written

$$\begin{aligned} \frac{z_1}{z_2} &= \frac{r_1(\cos\theta_1 + i\sin\theta_1)}{r_2(\cos\theta_2 + i\sin\theta_2)} \\ &= \frac{r_1(\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 - i\sin\theta_2)}{r_2(\cos\theta_2 + i\sin\theta_2)(\cos\theta_2 - i\sin\theta_2)} \\ &= \frac{r_1}{r_2}\,\frac{(\cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2) + i(\sin\theta_1\cos\theta_2 - \cos\theta_1\sin\theta_2)}{\cos^2\theta_2 + \sin^2\theta_2} \\ &= \frac{r_1}{r_2}[\cos(\theta_1-\theta_2) + i\sin(\theta_1-\theta_2)] \qquad r_2 \neq 0 \qquad (5) \end{aligned}$$

In words, then, *the product of two complex numbers is a complex number whose absolute value is the product of the absolute values of the two factors and whose amplitude is the sum of the amplitudes of the two factors*, and *the quotient of two complex numbers is a complex number whose absolute value is the quotient of the absolute values of the numbers and whose amplitude is the difference of their amplitudes.* The behavior of the angles of complex numbers when the numbers are multiplied or divided is concisely expressed by the formulas

(6) $$\arg z_1z_2 = \arg z_1 + \arg z_2$$

(7) $$\arg \frac{z_1}{z_2} = \arg z_1 - \arg z_2$$

In Sec 14 7, when we succeed in writing a general complex number as an exponential, the reason for the striking resemblance of these results to the corresponding logarithmic formulas will become apparent.

The extension of these ideas to products of more than two factors is obvious, and we can write at once

$$z_1z_2 \cdots z_n = r_1r_2 \cdots r_n[\cos(\theta_1 + \theta_2 + \cdots + \theta_n) + i\sin(\theta_1 + \theta_2 + \cdots + \theta_n)]$$

In particular, if all the z's are the same, we have the important result

(8) $$z^n = r^n(\cos n\theta + i\sin n\theta)$$

If $r = 1$, this is known as **de Moivre's theorem.*** Since the law of division in polar form (5) gives

$$\frac{1}{z} = \frac{1}{r}[\cos(0 - \theta) + i\sin(0 - \theta)] = \frac{1}{r}[\cos(-\theta) + i\sin(-\theta)]$$

which is just the content of Eq (8) for $n = -1$, it is clear that this formula is valid for all integral values of n, both positive and negative

The extension of Eq (8) to roots of integral order is an easy matter In fact, an nth root of $z = r(\cos\theta + i\sin\theta)$ is defined as any number $w = R(\cos\phi + i\sin\phi)$ such that

$$w^n \equiv R^n(\cos n\phi + i\sin n\phi) = z = r(\cos\theta + i\sin\theta)$$

Since two complex numbers which are equal must have the same modulus, it follows that

$$R^n = r \qquad \text{or} \qquad R = r^{1/n}$$

It should be noted that only real numbers are involved in the determination of R, since $r^{1/n}$ is the *real* nth root of the positive quantity r and can be found by an ordinary logarithmic calculation Also, the angles of equal complex numbers must either be equal or differ at most by an integral multiple of 2π Hence,

$$n\phi = \theta + 2k\pi \qquad \text{or} \qquad \phi = \frac{\theta + 2k\pi}{n}$$

For $k = 0, 1, \ldots, n - 1$, these values of ϕ define distinct angles; after this the same angles are repeated, again and again, each time with an irrelevant increment of 2π in their measures

---

* Named for the French mathematician Abraham de Moivre (1667–1754), although an equivalent form had been obtained earlier by the Englishman Roger Cotes (1682–1716)

Thus *there are exactly n distinct values of* $w = z^{1/n}$

$$w = z^{1/n} = r^{1/n}\left(\cos\frac{\theta + 2k\pi}{n} + i\sin\frac{\theta + 2k\pi}{n}\right) \qquad k = 0, 1, \ldots, n - 1 \tag{9}$$

In the complex plane these are represented by radii of the circle with center at the origin and radius $r^{1/n}$, spaced at equal angular intervals of $2\pi/n$ from the radius whose angle is θ/n.

With integral powers and roots defined, the general rational power of a complex number can be defined at once In fact,

$$\begin{aligned} z^{p/q} = (z^{1/q})^p &= \left[r^{1/q}\left(\cos\frac{\theta + 2k\pi}{q} + i\sin\frac{\theta + 2k\pi}{q}\right)\right]^p \\ &= r^{p/q}\left[\cos\frac{p}{q}(\theta + 2k\pi) + i\sin\frac{p}{q}(\theta + 2k\pi)\right] \qquad k = 0, 1, \ldots, q - 1 \end{aligned} \tag{10}$$

The definition of z^α when α is not a rational number, however, must be postponed until Sec. 14.7.

EXAMPLE 1

Find the four fourth roots of $-8i$

To do this, we must first write $-8i$ in standard polar form

$$-8i = 8\left(\cos\frac{3\pi}{2} + i\sin\frac{3\pi}{2}\right)$$

From this, by applying Eq (9), we find that the four fourth roots of $-8i$ are given by the expression

$$8^{1/4}\left[\cos\frac{1}{4}\left(\frac{3\pi}{2} + 2k\pi\right) + i\sin\frac{1}{4}\left(\frac{3\pi}{2} + 2k\pi\right)\right] \qquad k = 0, 1, 2, 3$$

or, explicitly,

$$\begin{aligned} r_1 &= 8^{1/4}\left(\cos\frac{3\pi}{8} + i\sin\frac{3\pi}{8}\right) && k = 0 \\ r_2 &= 8^{1/4}\left(\cos\frac{7\pi}{8} + i\sin\frac{7\pi}{8}\right) && k = 1 \\ r_3 &= 8^{1/4}\left(\cos\frac{11\pi}{8} + i\sin\frac{11\pi}{8}\right) && k = 2 \\ r_4 &= 8^{1/4}\left(\cos\frac{15\pi}{8} + i\sin\frac{15\pi}{8}\right) && k = 3 \end{aligned}$$

The coefficient $8^{1/4}$ is, of course, the *real* fourth root of 8, the value of which is found by a simple logarithmic calculation to be 1.682

EXAMPLE 2

Using de Moivre's theorem and the binomial expansion, express $\cos 4\theta$ and $\sin 4\theta$ in terms of powers of $\cos\theta$ and $\sin\theta$

To do this we consider $(\cos\theta + i\sin\theta)^4$ and expand it first by de Moivre's theorem and then by the binomial theorem. This gives the identity

$$\begin{aligned} \cos 4\theta + i\sin 4\theta &= \cos^4\theta + 4i\cos^3\theta\sin\theta + 6i^2\cos^2\theta\sin^2\theta + 4i^3\cos\theta\sin^3\theta + i^4\sin^4\theta \\ &= (\cos^4\theta - 6\cos^2\theta\sin^2\theta + \sin^4\theta) + i(4\cos^3\theta\sin\theta - 4\cos\theta\sin^3\theta) \end{aligned}$$

Equating real and imaginary parts of these equal complex expressions, we obtain the required formulas

$$\cos 4\theta = \cos^4\theta - 6\cos^2\theta\sin^2\theta + \sin^4\theta$$

$$\sin 4\theta = 4(\cos^3\theta\sin\theta - \cos\theta\sin^3\theta)$$

EXERCISES

1 Show that multiplying a complex number by i rotates it through 90° without changing its length What is the effect of multiplying a complex number (a) by $-i$? (b) by $\sqrt{i}$?

2 A square lies entirely in the second quadrant If one of its sides joins the points -3 and $2i$, find the coordinates of the other two vertices

3 Find all the distinct fourth roots of -1.

4 Find all the distinct fifth roots of 32

5 Find all the distinct cube roots of i

6 Express the complex number $8 - 8\sqrt{3}\,i$ in polar form, and find its distinct fourth roots

7 Find the distinct cube roots of $1 + i$, and reduce each to the form $a + ib$, where a and b are decimal fractions

8 Find all the distinct values of $(1 - i)^{3/4}$

9 Find all the distinct values of $(-1 - i)^{3/5}$

10 Using de Moivre's theorem, express $\cos 5\theta$ and $\sin 5\theta$ in terms of powers of $\cos\theta$ and $\sin\theta$

11 Show that, if n is an integer, both $\cos n\theta$ and $(\sin n\theta)/(\sin\theta)$ can be expressed as polynomials in $\cos\theta$

12 If z_1 and z_2 are complex numbers, what point is represented by $(z_1 + z_2)/2$? What is the locus of the points $\lambda z_1 + \mu z_2$, where λ and μ are real parameters and $\lambda + \mu = 1$?

13 Show that the centroid of a system of three particles of equal mass situated at the points z_1, z_2, z_3 is the point $(z_1 + z_2 + z_3)/3$ Where is the centroid of a system of three masses m_1, m_2, and m_3 situated respectively at the points z_1, z_2, and z_3?

14 Using the polar form of the multiplication law, devise a geometrical construction for the product of two complex numbers

15 Devise a geometrical construction for the quotient of two complex numbers

14.4

Absolute values

We have already defined the absolute value of a complex number z as the length of its representative vector; i e.,

$$|z| = \sqrt{x^2 + y^2} = \sqrt{\mathcal{R}^2(z) + \mathcal{I}^2(z)}$$

From this it is evident that *a complex number is zero if and only if its absolute value is zero* Since $\mathcal{R}^2(z)$ and $\mathcal{I}^2(z)$ are both nonnegative real numbers, it is also clear, dropping first one and then the other of these quantities from the last equation, that*

(1) $$|z| \geqq \mathcal{R}(z)$$

(2) $$|z| \geqq \mathcal{I}(z)$$

---

* We must always keep in mind the fact that the complex numbers cannot be ordered and that *greater than* and *less than* have meaning only when applied to real numbers.

Moreover, from the definition of conjugate complex numbers, it follows that

$$|z| = |\bar{z}| \tag{3}$$

and

$$z \cdot \bar{z} = |z|^2 \tag{4}$$

Also, from Eqs. (4) and (5), Sec 14.3, for the products and quotients of complex numbers expressed in polar form, it is clear that

$$|z_1 z_2| = |z_1| \cdot |z_2| \tag{5}$$

and

$$\left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|} \qquad z_2 \neq 0 \tag{6}$$

Since the length of any side of a triangle must be equal to or less than the sum of the lengths of the other two sides, it follows from the geometric addition of complex numbers (Fig. 14 4a) that

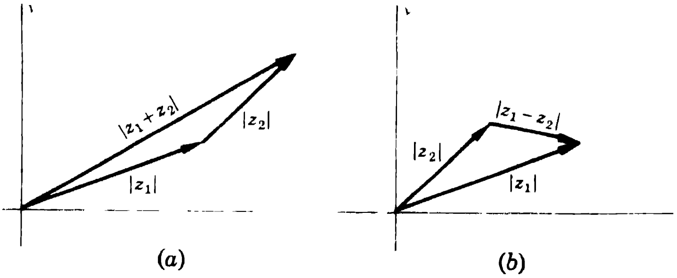


FIGURE 14.4 The triangle inequality applied to complex numbers

$$|z_1 + z_2| \leqq |z_1| + |z_2| \tag{7}$$

This can readily be extended to three terms, for

$$\begin{aligned} |z_1 + z_2 + z_3| &= |z_1 + (z_2 + z_3)| \\ &\leqq |z_1| + |z_2 + z_3| \\ &\leqq |z_1| + |z_2| + |z_3| \end{aligned}$$

The important extension to n terms is obvious:

$$\left|\sum_{k=1}^{n} z_k\right| \leqq \sum_{k=1}^{n} |z_k| \tag{8}$$

It is also geometrically evident that the length of any side of a triangle must be at least as great as the difference of the lengths of the other two sides (Fig. 14.4b). Hence,

$$|z_1 - z_2| \geqq \big|\, |z_1| - |z_2| \,\big| \geqq 0 \tag{9}$$

If it happens that $|z_1|$ is greater than or equal to $|z_2|$, the outer absolute-value signs on the right are, of course, unnecessary.

EXAMPLE 1

Describe the region in the z-plane defined by the inequality $\mathcal{R}(z) > 1$

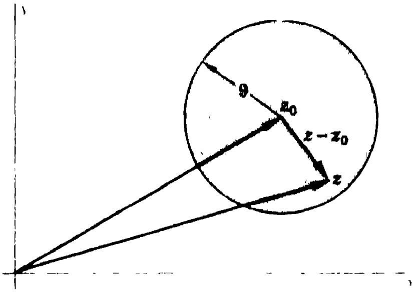
If the real part of z is greater than 1, the image of z must be a point to the right of the line $x = 1$ Hence, the given inequality defines the set of all points in the half plane to the right of this line Since the equality sign is not included in the definition of the region, points actually on the line $x = 1$ do not belong to the region

EXAMPLE 2

What region in the z-plane is defined by $|z - z_0| \leq 9$?

In words, the given inequality asserts that the distance between the image point of z and the fixed point which is the image of z_0 is equal to or less than 9 This clearly defines the set of all points within and on the boundary of the circle of radius 9 which has the image of z_0 as its center (Fig 14 5) In the work that lies ahead, we shall frequently have to consider regions of this type

FIGURE 14.5
The circular region
$|z - z_0| \leq 9$



EXAMPLE 3

If $w = (z + i)/(iz + 1)$, show that the restriction $\mathscr{I}(z) \leq 0$ implies the restriction $|w| \leq 1$

Since we are asked to establish a certain property of $|w|$, our first step is to compute this quantity This can be done in various ways, but it is probably most convenient to construct the product

$$w\,\bar{w} = |w|^2 = \frac{z+i}{iz+1}\cdot\overline{\left(\frac{z+i}{iz+1}\right)}$$

Since the conjugate of a quotient is the quotient of the conjugates, this can be written as

$$|w|^2 = \frac{z+i}{iz+1}\;\frac{\overline{z+i}}{\overline{iz+1}}$$

Moreover, the conjugate of a sum is the sum of the conjugates, hence we have further

$$|w|^2 = \frac{z+i}{iz+1}\cdot\frac{\bar{z}+\bar{i}}{\overline{iz}+1}$$

Finally, since $\bar{i} = -i$ and $\overline{iz} = \bar{i}\bar{z} = -i\bar{z}$, we have

$$|w|^2 = \frac{z+i}{iz+1}\;\frac{\bar{z}-i}{-i\bar{z}+1} = \frac{(z\bar{z}+1) - i(z-\bar{z})}{(z\bar{z}+1) + i(z-\bar{z})}$$

$$= \frac{z\bar{z}+1+2\mathscr{I}(z)}{z\bar{z}+1-2\mathscr{I}(z)}$$

Now, $z\bar{z} + 1$ is a positive quantity Hence, it is clear that, if $\mathscr{I}(z) \leq 0$, as given, then the numerator of the last fraction is equal to or less than the denominator Thus $|w|^2$, and, hence, $|w|$, is at most equal to 1 under the given conditions

Since the restriction $\mathcal{I}(z) \leqq 0$ implies that z lies on or below the real axis in the plane in which z is plotted and since $|w| \leqq 1$ implies that w lies on or inside the unit circle in the plane in which w is plotted, it follows that the given relation

$$w = \frac{z+i}{iz+1}$$

can be thought of as a transformation, or mapping, which sends the lower half of the z-plane, point by point, into the region consisting of the unit circle and its interior in the w-plane. Mappings of this sort are of considerable importance in applied mathematics, and in Chap. 17 we shall examine their properties in more detail.

EXERCISES

1 If a and b are real, show that $\left|\dfrac{a+ib}{b+ia}\right| = 1$. Is this true if a and b are not real?

2 Find $|z|$, $\mathcal{R}(z)$, and $\mathcal{I}(z)$ if $z = [(3+4i)(12-5i)]/2i$.

3 Under what conditions will $|z_1 + z_2| = |z_1| + |z_2|$?

4 Show that $\left|\dfrac{z_1}{z_1+z_2}\right| \leqq \dfrac{|z_1|}{\big||z_1| - |z_2|\big|}$. Under what conditions will the equality sign hold?

5 Show that $|x+iy| \geqq (|x|+|y|)/\sqrt{2}$. Under what conditions will the equality sign hold?

6 Show that $|z_1 - z_2|^2 + |z_1 + z_2|^2 = 2|z_1|^2 + 2|z_2|^2$.

7 Show that the locus of points for which $\left|\dfrac{z-1}{z+1}\right| = k$, where k is a positive constant different from 1, is a circle. What is the locus if $k = 1$? if $k = 0$? if $k < 0$?

8 What region in the z-plane is defined by the inequalities $0 < \mathcal{R}(z) \leqq \mathcal{I}(z)$?

9 What region in the z-plane is defined by the inequality $|z-1| \leqq \mathcal{R}(z)$?

10 If $w = i(1-z)/(1+z)$, prove that $|z| < 1$ implies $\mathcal{I}(w) > 0$.

11 Without using any properties of the polar representation of complex numbers, prove that $|z_1 z_2| = |z_1|\,|z_2|$.

12 Prove algebraically that $|z_1 + z_2| \leqq |z_1| + |z_2|$. [Hint: Consider the identity $|z_1+z_2|^2 = (z_1+z_2)\overline{(z_1+z_2)}$.]

13 Prove algebraically that $|z_1 - z_2| \geqq \big||z_1| - |z_2|\big| \geqq 0$.

14 Prove that, if $z + 1/z$ is real, then either z is real or the absolute value of z is 1.

15 If $z_1, z_2, \ldots, z_n$ and $w_1, w_2, \ldots, w_n$ are complex numbers, prove that

$$\left|\sum_{i=1}^{n} z_i w_i\right|^2 \leqq \sum_{i=1}^{n} |z_i|^2 \sum_{i=1}^{n} |w_i|^2$$

This result is sometimes known as **Cauchy's inequality**. [Hint: Consider the discriminant of the quadratic equation $\sum_{i=1}^{n} (|z_i|\lambda - |w_i|)^2 = 0$.]

14.5

Functions of a complex variable

If $z = x + iy$ and $w = u + iv$ are two complex variables, and if, for each value of z in some portion of the complex plane, one or more values of w are defined, then w is said to be a **function** of z, and we write

$$w = f(z)$$

If $w = f(z)$, that is, if

$$u + iv = f(x + iy)$$

it follows that the real numbers u and v are themselves determined by the real numbers x and y. Hence, the assertion that w is a function of $z = x + iy$ can also be written

(1) $$w = u(x,y) + iv(x,y)$$

where $u(x,y)$ and $v(x,y)$ are real-valued functions of the real variables x and y. Clearly, whenever a value of z is given, values of x and y are thereby provided, and, thus, one or more values of w are determined by (1). For example, if

$$w = f(z) = (x^2 - y) + (x + y^2)i$$

and if $z = 1 + 2i$, then $x = 1$ and $y = 2$, and, thus,

$$f(1 + 2i) = (1^2 - 2) + (1 + 2^2)i = -1 + 5i$$

If w is defined as a function of z in the form (1), it may be possible by suitable manipulations to rearrange the expression $u(x,y) + iv(x,y)$ so that x and y occur only in the binomial combination $x + iy$. For instance,

$$w = (x^2 - y^2) + 2ixy$$

is immediately recognizable as

$$w = (x + iy)^2 = z^2$$

and $$w = \frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2}$$

is nothing but the standard complex form of

$$w = \frac{1}{x + iy} = \frac{1}{z}$$

On the other hand, it may be impossible to express w in a form involving only the explicit combination $x + iy$ without using such "artificial" expressions as $\mathcal{R}(z) \equiv x$ and $\mathcal{I}(z) \equiv y$, with which, of course, any formula in x and y can be written in terms of z. For instance, unless we resort to "artificial" functions, no rearrangement of the formula

$$w = 7x + 3iy = 4\mathcal{R}(z) + 3z = 7z - 4i\mathcal{I}(z) = 5z + 2\bar{z}$$

can reduce w to explicit dependence on z alone. In our work and, in fact, in most applications of complex variable theory, the only functions of interest will be those which can be written in terms of z alone, without recourse to $\bar{z}$, $\mathcal{R}(z)$, $\mathcal{I}(z)$, and similar expressions.

Frequently our interest in a function will be restricted to its behavior at the points of some specified part of the z-plane.

However, before we can undertake discussions of this sort, we must define and explain some of the simpler properties of the sets of points we intend to consider

By a **neighborhood** of a point z_0 we mean any set consisting of all the points which satisfy an inequality of the form

$$|z - z_0| < \epsilon \qquad \epsilon > 0$$

Geometrically speaking, a neighborhood of z_0 thus consists of all the points within but not on a circle having z_0 as center A point z_0 belonging to a set S is said to be an **interior point** of S if there exists at least one neighborhood of z_0 whose points all belong to S A set each of whose points is an interior point is said to be **open.** A point z_0 not belonging to a set S is said to be **exterior** to S if there exists at least one neighborhood of z_0 none of whose points belongs to S. Intermediate between points interior to S and points exterior to S are the boundary points of S A point z_0 is said to be a **boundary point** of a set S if every neighborhood of z_0 contains both points belonging to S and points not belonging to S. The boundary points of a set may or may not belong to the set, depending upon its definition

A point z_0 is said to be a **limit point** of a set if every neighborhood of the point contains at least one point of the set distinct from z_0 A set which contains all its limit points is said to be **closed.** Obviously, a set can be defined to contain some but not all of its limit points, hence it is clear that a set may be neither open nor closed

If a set S has the property that every pair of its points can be joined by a polygonal arc whose points all belong to the set, it is said to be **connected.** An open connected set is said to be a **region** or a **domain.** A set consisting of a region together with all its limit points is called a **closed region.** A connected set S with the property that every simple closed curve* which can be drawn in its interior encloses only points of S is said to be **simply connected.** If it is possible to draw in S at least one simple closed curve whose interior contains points not belonging to S, then S is said to be **multiply connected.**† If there exists a circle with center at the origin enclosing all the points of a set S, i e , if there exists a number d such that

$$|z| < d \qquad \text{for all } z \text{ in } S$$

then S is said to be **bounded.** A set which is not bounded is said to be **unbounded.** The set consisting of the points between two concentric circles is called an **annular region** or an **annulus.**

---

* See footnote to Theorem 1, Sec 12 4

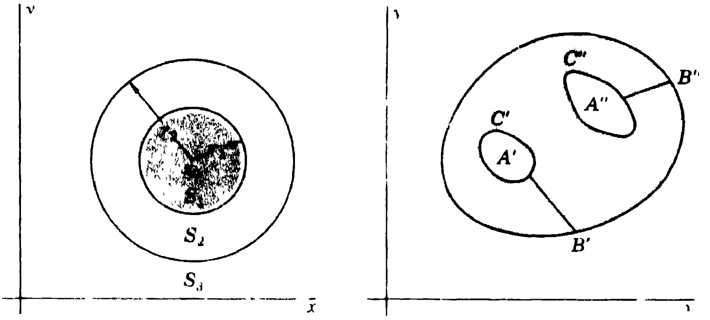
† In two dimensions the definitions of simply connected sets and multiply connected sets given at the end of Sec 12 5 are clearly equivalent to those of the present section.

The preceding ideas are illustrated in Fig. 14.6*a*, where the three sets

S_1 $\quad |z - z_0| < r_1$

S_2: $\quad r_1 \leqq |z - z_0| < r_2$

S_3: $\quad r_2 \leqq |z - z_0|$



(*a*) (*b*)

FIGURE 14.6 Typical regions in the complex plane.

are shown. The set S_1 consists of all points interior to the circle $|z - z_0| = r_1$. It is bounded and simply connected. Since points on the boundary circle $|z - z_0| = r_1$ are not included in the definition of S_1, the set is open and is, therefore, a domain, in particular it is a neighborhood of z_0. The set S_2 consists of all the points in the annulus between the circles $|z - z_0| = r_1$ and $|z - z_0| = r_2$ plus the points on the inner boundary of the annulus but not those on the outer boundary. Since S_2 thus contains some but not all of its boundary points, it is neither open nor closed and is, therefore, neither a domain nor a closed region. Clearly, there are closed curves in S_2, namely, any curve encircling the inner boundary, which will enclose points not belonging to S_2, namely, the points of S_1. Hence, S_2 is multiply connected. Obviously, S_2 is bounded. The set S_3 consists of all points on and outside the circle $|z - z_0| = r_2$. It is, therefore, unbounded, closed, and multiply connected. According to our definition, it is a closed region.

Because simply connected regions are in many respects easier to work with than multiply connected regions, it is often desirable to be able to reduce the latter to the former. This can always be done by modifying the given multiply connected region through the introduction of auxiliary boundary arcs, or **crosscuts,** joining boundary curves that were originally disconnected. The effectiveness of this technique is illustrated in Fig. 14.6*b*, which shows a closed region originally multiply connected with one outer boundary curve C and two inner boundary curves C' and C''. The introduction of the auxiliary boundary arcs $A'B'$ and $A''B''$ clearly makes it impossible to draw closed curves which lie entirely in the *interior* of the modified region and at the same time encircle either of the inner boundaries C' and C''. The modified region is, therefore, simply connected, as required.

It will often be necessary for us to consider the **limit of a function of z** as z approaches some particular value z_0. The basis for this is the following definition

DEFINITION 1

If $f(z)$ is a single-valued function of z and w_0 is a complex constant and if, for every $\epsilon > 0$, there exists a positive number $\delta(\epsilon)$ such that $|f(z) - w_0| < \epsilon$ for all z such that $0 < |z - z_0| < \delta$, then w_0 is said to be the limit of $f(z)$ as z approaches z_0

In less technical terms, w_0 is the limit of $f(z)$ as z approaches z_0 provided that $f(z)$ can be kept arbitrarily close to w_0 by keeping z sufficiently close to but distinct from z_0

EXAMPLE 1

If $f(z) = (x + y)^2/(x^2 + y^2)$, show that

$$\lim_{x\to 0}\,[\lim_{y\to 0} f(z)] = 1 \qquad \text{and} \qquad \lim_{y\to 0}\,[\lim_{x\to 0} f(z)] = 1$$

but that $\lim_{z\to 0} f(z)$ does not exist

Clearly,
$$\lim_{x\to 0}\,[\lim_{y\to 0} f(z)] = \lim_{x\to 0}\left[\lim_{y\to 0}\frac{(x+y)^2}{x^2+y^2}\right] = \lim_{x\to 0}\,(1) = 1$$

and
$$\lim_{y\to 0}\,[\lim_{x\to 0} f(z)] = \lim_{y\to 0}\left[\lim_{x\to 0}\frac{(x+y)^2}{x^2+y^2}\right] = \lim_{y\to 0}\,(1) = 1 \qquad \text{as asserted}$$

On the other hand, for $\lim_{z\to 0} f(z)$ to exist, it is necessary that $f(z)$ approach the same value along all paths leading to the origin, and this is not the case; for along the paths $y = mx$ we have

$$\lim_{z\to 0} f(z) = \lim_{z\to 0}\frac{(x+y)^2}{x^2+y^2} = \lim_{x\to 0}\frac{(1+m)^2}{1+m^2} = \frac{(1+m)^2}{1+m^2}$$

The limiting value here clearly depends on m, that is, $f(z)$ approaches different values along different radial lines, and hence no limit exists

Closely associated with the concept of a limit is the concept of **continuity**:

DEFINITION 2

The function $f(z)$ is continuous at the point z_0 provided that $\lim_{z\to z_0} f(z) = f(z_0)$

In other words, for a function to be continuous at a point z_0, the function must have both a value at that point and a limit as z approaches that point, and the two must be equal If $f(z)$ is continuous at every point of a region, it is said to be **continuous throughout the region.**

In addition to the fundamental theorems on limits we encountered in calculus, there are various theorems on continuous functions which we shall need from time to time. For the most part these appear almost self-evident, although their proofs are by no means trivial. We shall merely list them here, and refer to standard texts on advanced calculus for their proof.*

THEOREM 1

Sums, differences, and products of continuous functions and quotients of continuous functions, provided the divisor functions are different from zero, are continuous.

THEOREM 2

A continuous function of a continuous function is continuous.

THEOREM 3

A necessary and sufficient condition that

$$f(z) = u(x,y) + iv(x,y)$$

be continuous is that the real functions $u(x,y)$ and $v(x,y)$ be continuous

THEOREM 4

If $f(z)$ is continuous at a point z_0 and if $f(z_0) \neq 0$, then there exists a neighborhood of z_0 throughout which $f(z)$ is different from 0.

THEOREM 5

If $f(z)$ continuous over a bounded, closed region R, then there exists a positive constant M such that $|f(z)| < M$ for all values of z in R

EXERCISES

1 If $f(z) = xy + i(x^2 - y^2)$, what is $f(-1 + 2i)$?

2 If $f(z) = z + (\bar{z})^2 + \mathcal{I}(z\bar{z})$, what is $f(2 + i)$?

3 Express $(2xy + 2x - 1) - i(x^2 - y^2 - 2y)$ as a polynomial in the binomial argument $z = x + iy$

4 Express $x^2 + iy^2$ in terms of z and $\bar{z}$

5 Describe each of the following sets of points, telling whether it is bounded or unbounded, open or closed, and simply or multiply connected

a $\mathcal{I}(z) > 0$
b $2 \leqq |z| \leqq 3$
c $|z - 1| > 4$
d $0 \leqq \mathcal{R}(z) \leqq 1$
e $0 \leqq \mathcal{I}(z) < \mathcal{R}(z)$
f $|z^2 - 1| \leqq 5/4$

6 Show that $\lim\limits_{z\to 0} \dfrac{xy}{x^2 + y^2}$ does not exist

7 Show that $\lim\limits_{z\to 0} \dfrac{x^2y}{x^4 + y^2}$ does not exist, even though this function approaches the same limit along every straight line through the origin

8 If $f(z) = \begin{cases} x \sin 1/y & y \neq 0 \\ 0 & y = 0 \end{cases}$ show that $\lim\limits_{y\to 0} [\lim\limits_{x\to 0} f(z)]$ and $\lim\limits_{z\to 0} f(z)$ exist and are equal, but that $\lim\limits_{x\to 0} [\lim\limits_{y\to 0} f(z)]$ does not exist

9 Prove Theorem 2

10 Show that every neighborhood of a limit point of a set S contains infinitely many points of S

---

* See, for instance, A E Taylor, "Advanced Calculus," pp 494–503, Ginn and Company, Boston, 1955

14.6

Analytic functions

The derivative of a function of a complex variable $w = f(z)$ is defined as

$$\frac{dw}{dz} = w' = f'(z) = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} \tag{1}$$

This definition is formally identical with that for the derivative of a function of a real variable. Moreover, since the general theory of limits is phrased in terms of absolute values, it is valid for complex variables as well as for real variables. Hence it is clear that formulas for the differentiation of functions of a real variable will have identical counterparts in the field of complex numbers when the corresponding functions of a complex variable are suitably defined. In particular, such familiar formulas as

$$\frac{d(w_1 \pm w_2)}{dz} = \frac{dw_1}{dz} \pm \frac{dw_2}{dz}$$

$$\frac{d(w_1 w_2)}{dz} = w_1 \frac{dw_2}{dz} + w_2 \frac{dw_1}{dz}$$

$$\frac{d(w_1/w_2)}{dz} = \frac{w_2(dw_1/dz) - w_1(dw_2/dz)}{w_2^2} \qquad w_2 \neq 0$$

$$\frac{d(w^n)}{dz} = nw^{n-1}\frac{dw}{dz}$$

are valid when w_1, w_2, and w are differentiable functions of a complex variable z. However, $\Delta z = \Delta x + i\,\Delta y$ is itself a complex variable, and the question of just how it is to approach zero involves difficulties which have no counterpart in the differentiation of functions of a real variable.

In Fig. 14.7, it is clear that Δz can approach zero; i.e., that a point

$Q\colon z + \Delta z$

can approach the point $P{:}z$, along infinitely many paths. In

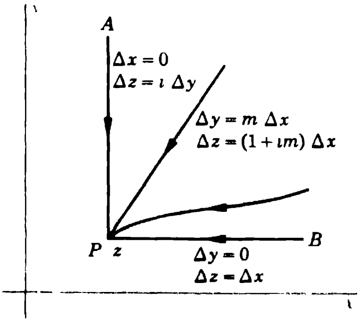


FIGURE 14.7 Plot showing various ways in which Δz can approach zero

particular, Q can approach P along the line AP on which Δx is zero or along the line BP on which Δy is zero. Clearly, *for the derivative of $f(z)$ to exist, it is necessary that the limit of the difference quotient* (1) *be the same no matter how Δz approaches zero* How severe a restriction this is can be seen by considering the simple function

$$w = f(z) = \bar{z} = x - iy$$

Giving to z the increment $\Delta z = \Delta x + i\,\Delta y$ means that x changes by the amount Δx and y changes by the amount Δy Hence,

$$\frac{f(z + \Delta z) - f(z)}{\Delta z} = \frac{[(x + \Delta x) - i(y + \Delta y)] - (x - iy)}{\Delta x + i\,\Delta y} = \frac{\Delta x - i\,\Delta y}{\Delta x + i\,\Delta y}$$

Now, if Δz is real, so that $\Delta y = 0$, we have

$$\lim_{\Delta z \to 0} \frac{\Delta x - i\,\Delta y}{\Delta x + i\,\Delta y} = \lim_{\Delta x \to 0} \frac{\Delta x}{\Delta x} = 1$$

On the other hand, if Δz is imaginary, so that $\Delta x = 0$, we have

$$\lim_{\Delta z \to 0} \frac{\Delta x - i\,\Delta y}{\Delta x + i\,\Delta y} = \lim_{\Delta y \to 0} \frac{-i\,\Delta y}{i\,\Delta y} = -1$$

More generally, if we let $\Delta z \to 0$ in such a way that $\Delta y = m\,\Delta x$, we have

$$\lim_{\Delta z \to 0} \frac{\Delta x - i\,\Delta y}{\Delta x + i\,\Delta y} = \lim_{\Delta x \to 0} \frac{\Delta x - im\,\Delta x}{\Delta x + im\,\Delta x} = \frac{1 - im}{1 + im} = \frac{(1 - m^2) - 2im}{1 + m^2}$$

Thus, there are infinitely many complex values which the difference quotient for $f(z) = x - iy$ can be made to approach by choosing properly the manner in which Δz shall approach zero It is, therefore, apparent that $\bar{z} = x - iy$ has no derivative

That a function as simple as $f(z) = x - iy$ should have no derivative seems at first glance a discouraging state of affairs. However, there are many functions of z which do have derivatives, and in applications it is these functions which are of importance. Our immediate task is to identify these functions by obtaining conditions for the existence of the derivative of a function of a complex variable.

To do this, consider

$$w = f(z) = u(x,y) + w(x,y)$$

By definition,

$$(2)\quad \frac{dw}{dz} = \lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z}$$

$$= \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{[u(x + \Delta x, y + \Delta y) + iv(x + \Delta x, y + \Delta y)] - [u(x,y) + iv(x,y)]}{\Delta x + i\,\Delta y}$$

Now, if Δz is real, i.e., if $\Delta y = 0$, we obtain

$$\frac{dw}{dz} = \lim_{\Delta x \to 0} \frac{[u(x + \Delta x, y) + iv(x + \Delta x, y)] - [u(x,y) + iv(x,y)]}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \left[\frac{u(x + \Delta x, y) - u(x,y)}{\Delta x} + i \frac{v(x + \Delta x, y) - v(x,y)}{\Delta x} \right]$$

The two difference quotients which appear in the last expression are precisely those whose limits define the partial derivatives of u and v with respect to x. Hence, it appears that

(3) $$\frac{dw}{dz} = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x}$$

On the other hand, if Δz is imaginary, i.e., if $\Delta x = 0$, we find from (2) that

$$\frac{dw}{dz} = \lim_{\Delta y \to 0} \frac{[u(x, y + \Delta y) + iv(x, y + \Delta y)] - [u(x,y) + iv(x,y)]}{i\,\Delta y}$$

$$= \lim_{\Delta y \to 0} \left[\frac{u(x, y + \Delta y) - u(x,y)}{i\,\Delta y} + i \frac{v(x, y + \Delta y) - v(x,y)}{i\,\Delta y} \right]$$

$$= \frac{1}{i} \frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}$$

or, finally,

(4) $$\frac{dw}{dz} = \frac{\partial v}{\partial y} - i \frac{\partial u}{\partial y}$$

Thus, if the derivative dw/dz is to exist, it is necessary that the two expressions we have just derived for it be the same. Hence, from (3) and (4),

$$\frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i \frac{\partial u}{\partial y}$$

which requires that

(5a) $$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$$

(5b) $$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

These two extremely important conditions, which are known as the **Cauchy-Riemann equations,*** have arisen here from a consideration of only two of the infinitely many ways in which Δz can approach zero. It is, therefore, natural to expect that severe additional conditions will be necessary to ensure that along these other paths $\Delta w/\Delta z$ will also approach the same limit dw/dz. This is not the case, however, and it can be proved without great difficulty* that, if u and v together with their first partial derivatives u_x, u_y, v_x, v_y are continuous in some neighborhood of the point z_0, then the Cauchy-Riemann equations are not only necessary but also sufficient conditions for the existence of a derivative of $w = u(x,y) + iv(x,y)$ at $z = z_0$

---

* After the French mathematician Augustin Louis Cauchy (1789–1857), and the German mathematician George Friedrich Bernhard Riemann (1826–1866).

If $w = f(z)$ possesses a derivative at $z = z_0$ and at every point in some neighborhood of z_0, then $f(z)$ is said to be **analytic** at z_0, and z_0 is called a **regular point** of the function If $f(z)$ is not analytic at z_0, but if every neighborhood of z_0 contains points at which $f(z)$ is analytic, then z_0 is called a **singular point** of $f(z)$ A function analytic at every point of a region R we shall call **analytic in** R Although most writers use this term, a few substitute such adjectives as **regular** and **holomorphic.** As a summary of our discussion we have the following theorem

THEOREM 1

If u and v are real single-valued functions of x and y which, with their four first partial derivatives, are continuous throughout a region R, then the Cauchy-Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \text{and} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

are both necessary and sufficient conditions that $f(z) = u(x,y) + iv(x,y)$ be analytic in R In this case the derivative of $f(z)$ is given by either of the expressions

$$f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} \qquad f'(z) = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y}$$

EXAMPLE 1

For $w = \bar{z} = x - iy$, we have $u = x$ and $v = -y$ In this case

$$\frac{\partial u}{\partial x} = 1 \qquad \frac{\partial u}{\partial y} = 0 \qquad \frac{\partial v}{\partial x} = 0 \qquad \frac{\partial v}{\partial y} = -1$$

and, although the second of the Cauchy-Riemann equations is satisfied everywhere, the first is nowhere satisfied Hence, there is no point in the z-plane where dw/dz exists, which, of course, confirms our earlier investigation of this function

EXAMPLE 2

For $w = z\bar{z} = x^2 + y^2$, we have $u = x^2 + y^2$ and $v = 0$ In this case the partial derivatives

$$\frac{\partial u}{\partial x} = 2x \qquad \frac{\partial u}{\partial y} = 2y \qquad \frac{\partial v}{\partial x} = 0 \qquad \frac{\partial v}{\partial y} = 0$$

are continuous everywhere. However, the Cauchy-Riemann equations, which in this case are, respectively,

$$2x = 0 \qquad \text{and} \qquad 2y = 0$$

are satisfied only at the origin Hence, $z = 0$ is the only point at which dw/dz exists, and $w = z\bar{z}$ is nowhere analytic

---

* See, for instance, Einar Hille, "Analytic Function Theory," vol 1, pp 78–80, Ginn and Company, Boston, 1959

EXAMPLE 3

For $w = z^2 = (x^2 - y^2) + 2ixy$, we have

$$\frac{\partial u}{\partial x} = 2x \qquad \frac{\partial u}{\partial y} = -2y \qquad \frac{\partial v}{\partial x} = 2y \qquad \frac{\partial v}{\partial y} = 2x$$

and the Cauchy-Riemann equations are identically satisfied. Moreover, the first partial derivatives of u and v are everywhere continuous. Hence, the derivative dw/dz exists at all points of the z-plane, and its value from either (3) or (4) is

$$\frac{dw}{dz} = 2x + 2iy = 2z$$

This, of course, is exactly what formal differentiation according to the power rule would give.

Analytic functions have a great many important properties, many of which we shall investigate in later sections. At this point we note only the following.

PROPERTY 1

If both the real part and the imaginary part of an analytic function have continuous second partial derivatives, then they satisfy Laplace's equation

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0$$

PROOF Let $w = u(x,y) + iv(x,y)$ be an analytic function of z. Then u and v must satisfy the Cauchy-Riemann equations, namely,

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \text{and} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

If we differentiate the first of these with respect to x and the second with respect to y and add the results, we obtain the first assertion of the theorem

$$\begin{array}{rcl} \dfrac{\partial^2 u}{\partial x^2} &=& \dfrac{\partial^2 v}{\partial x\,\partial y} \\ \dfrac{\partial^2 u}{\partial y^2} &=& -\dfrac{\partial^2 v}{\partial y\,\partial x} \\ \hline \dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} &=& 0 \end{array}$$

The existence of the second partial derivatives and their continuity, which makes the order of differentiation in the cross partial derivatives immaterial, must here be assumed. Later we shall show that an analytic function possesses not only a first derivative, but derivatives of *all* orders, which implies the existence and continuity of all the partial derivatives of u and v. In exactly the same way it can be shown that v satisfies Laplace's equation.

A function which possesses continuous second partial derivatives and satisfies Laplace's equation is usually called a **harmonic function.** Two harmonic functions u and v so related that $u + iv$ is an analytic function are called **conjugate harmonic functions.*** This use of the word *conjugate* must not be confused with its use in describing $\bar{z}$, the complex number conjugate to z.

---

* The order in the pair (u,v) is important, as Exercise 6 makes clear.

PROPERTY 2

If $w = u(x,y) + iv(x,y)$ is an analytic function of z, then the curves of the family $u(x,y) = c$ are the orthogonal trajectories of the curves of the family $v(x,y) = k$, and vice versa

PROOF To prove this, we compute the slope of the general curve of each family by implicit differentiation, getting for the curves $u(x,y) = c$ the expression

$$\frac{dy}{dx} = -\frac{\partial u/\partial x}{\partial u/\partial y} \tag{6}$$

and for the curves $v(x,y) = k$ the expression

$$\frac{dy}{dx} = -\frac{\partial v/\partial x}{\partial v/\partial y} \tag{7}$$

By hypothesis, $w = u + iv$ is an analytic function Hence, it follows from Theorem 1 that u and v satisfy the Cauchy-Riemann equations Therefore, using these equations, the expression (7) for the slope of the general curve of the family $v(x,y) = k$ can be rewritten

$$\frac{dy}{dx} = \frac{\partial u/\partial y}{\partial u/\partial x}$$

which, at any common point, is just the negative reciprocal of the slope of the general curve of the family $u(x,y) = c$, as given by Eq (6) This suffices to prove that the two families of curves are orthogonal trajectories, as asserted

PROPERTY 3

If, in any analytic function $w = u(x,y) + iv(x,y)$, the variables x and y are replaced by their equivalents in terms of z and $\bar{z}$, namely,

$$x = \frac{z + \bar{z}}{2} \quad \text{and} \quad y = \frac{z - \bar{z}}{2i}$$

then w will appear as a function of z alone

PROOF Although z and $\bar{z}$ are clearly dependent, since either is determined when the other is given, we can regard w, by virtue of the given substitutions, as formally a function of two new independent variables z and $\bar{z}$ To show that w depends only on z and does not involve $\bar{z}$, it is sufficient to compute $\dfrac{\partial w}{\partial \bar{z}}$ and verify that it is identically zero Now,

$$\frac{\partial w}{\partial \bar{z}} = \frac{\partial(u + iv)}{\partial \bar{z}} = \frac{\partial u}{\partial \bar{z}} + i\frac{\partial v}{\partial \bar{z}} = \left(\frac{\partial u}{\partial x}\frac{\partial x}{\partial \bar{z}} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial \bar{z}}\right) + i\left(\frac{\partial v}{\partial x}\frac{\partial x}{\partial \bar{z}} + \frac{\partial v}{\partial y}\frac{\partial y}{\partial \bar{z}}\right)$$

Moreover, from the equations expressing x and y in terms of z and $\bar{z}$, we have

$$\frac{\partial x}{\partial \bar{z}} = \frac{1}{2} \qquad \frac{\partial y}{\partial \bar{z}} = -\frac{1}{2i} = \frac{i}{2}$$

Hence, we can write

$$\frac{\partial w}{\partial \bar{z}} = \left(\frac{1}{2}\frac{\partial u}{\partial x} + \frac{i}{2}\frac{\partial u}{\partial y}\right) + i\left(\frac{1}{2}\frac{\partial v}{\partial x} + \frac{i}{2}\frac{\partial v}{\partial y}\right) = \frac{1}{2}\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + \frac{i}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)$$

Since w, by hypothesis, is an analytic function, u and v satisfy the Cauchy-

Riemann equations, and, therefore, each of the quantities in parentheses in the last expression vanishes Thus $\dfrac{\partial w}{\partial \bar{z}} \equiv 0$. Hence, w is independent of $\bar{z}$, that is, depends on x and y only through the combination $z = x + iy$

EXERCISES

1 At what points does $(z - 2)/[(z + 1)(z^2 + 1)]$ fail to be analytic?

2 Show that at no point in the z-plane does the derivative of $f(z) = \Re(z) = x$ exist Does this contradict the fact that according to the rules of calculus $dx/dx = 1$? Explain

3 Where are the Cauchy-Riemann equations satisfied for the function $f(z) = xy^2 + ix^2y$? Where does $f'(z)$ exist? Where is $f(z)$ analytic?

4 Verify by direct substitution that $\Re(z^3)$ and $\Im(z^3)$ satisfy Laplace's equation

5 If $u + iv$ is an analytic function, under what conditions, if any, will $v + iu$ be analytic?

6 If u and v are conjugate harmonic functions, show that v and $-u$ as well as $-v$ and u are also conjugate harmonic functions, but that v and u are not

7 Show that the various values approached by the difference quotient of $f(z) = \bar{z}$ as $\Delta z \to 0$ along the lines $y = mx$ all lie on a circle

8 Is the converse of Property 2 true? That is, if $u(x,y) = c$ and $v(x,y) = k$ are orthogonal trajectories, is $u + iv$ necessarily an analytic function?

9 Prove that, if $f'(z) \equiv 0$, then $f(z)$ is a constant

10 If in the function $f(z) = u + iv$ we take z in polar form, namely,

$$z = r(\cos\theta + i\sin\theta)$$

show that the Cauchy-Riemann equations become

$$\frac{\partial u}{\partial r} = \frac{1}{r}\frac{\partial v}{\partial \theta} \qquad \text{and} \qquad \frac{\partial v}{\partial r} = -\frac{1}{r}\frac{\partial u}{\partial \theta}$$

11 If w is an analytic function of $z = r(\cos\theta + i\sin\theta)$, show that

$$\frac{dw}{dz} = (\cos\theta - i\sin\theta)\frac{\partial w}{\partial r} = -\frac{\sin\theta + i\cos\theta}{r}\frac{\partial w}{\partial \theta}$$

12 If $f(z)$ is an analytic function, show that

a $\left(\dfrac{\partial}{\partial x}|f(z)|\right)^2 + \left(\dfrac{\partial}{\partial y}|f(z)|\right)^2 = |f'(z)|^2$

b $\left(\dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2}\right)|f(z)|^2 = 4|f'(z)|^2$

13 If $f(z)$ and $\overline{f(z)}$ are both analytic functions, show that $f(z)$ is a constant

14 If $f(z)$ is an analytic function for which $u^2 + v^2$ is a constant, show that $f(z)$ is a constant

15 Prove L'Hospital's rule for analytic functions If $f(z)$ and $g(z)$ are analytic functions in a region containing z_0, if $f(z_0) = g(z_0) = 0$, and if $g'(z_0) \neq 0$, then $\lim\limits_{z \to z_0} \dfrac{f(z)}{g(z)} = \dfrac{f'(z_0)}{g'(z_0)}$

14.7

The elementary functions of z

The exponential function e^z is of fundamental importance, not only for its own sake, but also as a basis for defining all the other elementary functions. In its definition we seek to preserve as many of the characteristic properties of the real exponential function e^x as possible Specifically, we desire that

a e^z shall be single-valued and analytic
b $de^z/dz = e^z$
c e^z shall reduce to e^x when $\mathscr{I}(z) = 0$

If we let

(1) $$e^z = u + iv$$

and recall from Eq (3), Sec 14 6, that the derivative of an analytic function can be written in the form

$$f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}$$

then, to satisfy condition **b**, we must have

$$\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = u + iv$$

Hence, equating real and imaginary parts,

(2) $$\frac{\partial u}{\partial x} = u$$

(3) $$\frac{\partial v}{\partial x} = v$$

Now, Eq (2) will be satisfied if

(4) $$u = e^x\phi(y)$$

where $\phi(y)$ is any function of y Moreover, since e^z is to be analytic (condition **a**), u and v must satisfy the Cauchy-Riemann equations, hence, using the second of these equations, Eq (3) can be written

(5) $$-\frac{\partial u}{\partial y} = v$$

Differentiating this with respect to y, we obtain

$$\frac{\partial^2 u}{\partial y^2} = -\frac{\partial v}{\partial y}$$

or, replacing $\frac{\partial v}{\partial y}$ by $\frac{\partial u}{\partial x}$ according to the first of the Cauchy-Riemann equations,

$$\frac{\partial^2 u}{\partial y^2} = -\frac{\partial u}{\partial x}$$

Finally, using (2), this becomes

$$\frac{\partial^2 u}{\partial y^2} = -u$$

which, on substituting $u = e^x\phi(y)$ from (4), reduces to

$$e^x\phi''(y) = -e^x\phi(y) \qquad \text{or} \qquad \phi''(y) = -\phi(y)$$

This is a simple linear differential equation whose solution can be written down at once:

$$\phi(y) = A \cos y + B \sin y$$

Hence, from (4),

$$u = e^x\phi(y) = e^x(A \cos y + B \sin y)$$

and, from (5),

$$v = -\frac{\partial u}{\partial y} = -e^x(-A \sin y + B \cos y)$$

Therefore, from (1),

$$e^z = u + iv = e^x[(A \cos y + B \sin y) + i(A \sin y - B \cos y)]$$

If this is to reduce to e^x when $y = 0$, as required by condition **c**, we must have

$$e^x = e^x(A - iB)$$

which will be true if and only if $A = 1$ and $B = 0$.

Thus we have been led inevitably to the conclusion that *if* there is a function of z satisfying the conditions **a**, **b**, and **c**, *then* it must be

(6) $$e^z = e^{x+iy} = e^x(\cos y + i \sin y)$$

That this expression does, indeed, meet our requirements can be checked immediately; hence, we adopt it as the definition of e^z.

It is important to note that the right-hand side of (6) is in standard polar form. Hence,

$$\text{mod } e^z \equiv |e^z| = e^x \qquad \text{and} \qquad \arg e^z = y$$

The possibility of writing any complex number in exponential form is now apparent, for, applying (6), with $x = 0$ and $y = \theta$, we have

(7) $$\cos\theta + i \sin\theta = e^{i\theta}$$

and thus

(8) $$r(\cos\theta + i \sin\theta) = re^{i\theta}$$

The fact that the angle, or argument, of a complex number is actually an exponent explains why the angles of complex numbers are added when the numbers are multiplied and subtracted when the numbers are divided, as we found in Sec. 14.3.

From the relation

$$e^{i\theta} = \cos\theta + i \sin\theta$$

and its obvious companion

$$e^{-i\theta} = \cos(-\theta) + i \sin(-\theta) = \cos\theta - i \sin\theta$$

we obtain, by addition and subtraction, the so-called **Euler formulas**

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \qquad \text{and} \qquad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

On the basis of these equations, we extend the definitions of the sine and cosine into the complex domain by the formulas

(9) $$\cos z = \frac{e^{iz} + e^{-iz}}{2}$$

(10) $$\sin z = \frac{e^{iz} - e^{-iz}}{2i}$$

From these definitions it is easy to establish the validity of such familiar formulas as

$$\cos^2 z + \sin^2 z = 1$$

$$\cos(z_1 \pm z_2) = \cos z_1 \cos z_2 \mp \sin z_1 \sin z_2$$

$$\sin(z_1 \pm z_2) = \sin z_1 \cos z_2 \pm \cos z_1 \sin z_2$$

$$\frac{d(\cos z)}{dz} = -\sin z$$

$$\frac{d(\sin z)}{dz} = \cos z$$

If we expand the exponentials in (9), we find

$$\begin{aligned}\cos z &= \frac{e^{i(x+iy)} + e^{-i(x+iy)}}{2}\\ &= \frac{e^{-y}e^{ix} + e^{y}e^{-ix}}{2}\\ &= \frac{e^{-y}(\cos x + i\sin x) + e^{y}(\cos x - i\sin x)}{2}\\ &= \cos x\,\frac{e^{y} + e^{-y}}{2} - i\sin x\,\frac{e^{y} - e^{-y}}{2}\end{aligned}$$

or, using the usual definitions of the hyperbolic functions of real variables,

(11) $$\cos z = \cos(x + iy) = \cos x \cosh y - i \sin x \sinh y$$

Similarly, it is easy to show that

(12) $$\sin z = \sin(x + iy) = \sin x \cosh y + i \cos x \sinh y$$

In particular, taking $x = 0$ in (11) and (12), we find

(13) $$\cos iy = \cosh y$$

(14) $$\sin iy = i \sinh y$$

The remaining trigonometric functions of z are defined in terms of $\cos z$ and $\sin z$ by means of the usual identities.

EXAMPLE 1

What is $\cos(1 + 2i)$?

By direct use of (11) we have

$$\begin{aligned}\cos(1 + 2i) &= \cos 1 \cosh 2 - i \sin 1 \sinh 2\\ &= (0.5403)(3.7622) - i(0.8415)(3.6269)\\ &= 2.033 - 3.052i\end{aligned}$$

EXAMPLE 2

Prove that the only values of z for which $\sin z = 0$ are the real values $z = 0, \pm\pi, \pm 2\pi, \ldots$

From (12), $\sin z = \sin x \cosh y + i \cos x \sinh y$. Hence, if $\sin z$ is to vanish, it is necessary that simultaneously

$$\sin x \cosh y = 0$$

$$\cos x \sinh y = 0$$

Since y is a real number, it follows from the familiar properties of the hyperbolic cosine that $\cosh y \geqq 1$. Hence, the first of these equations can hold only if $\sin x = 0$, that is, if

$$x = 0, \pm\pi, \pm 2\pi, \ldots$$

But for these values of x, $\cos x$, being either 1 or -1, can never vanish. For the second equation to hold, it is therefore necessary that $\sinh y = 0$. Since y is real, the familiar properties of the hyperbolic sine can be invoked, leading to the conclusion that

$$y = 0$$

Hence, the only values of z for which $\sin z = 0$ are of the form

$$z = n\pi + 0i = n\pi \qquad n = 0, \pm 1, \pm 2, \ldots$$

The hyperbolic functions of z we define simply by extending the familiar definitions into the complex number field:

$$\cosh z = \frac{e^z + e^{-z}}{2} \tag{15}$$

$$\sinh z = \frac{e^z - e^{-z}}{2} \tag{16}$$

By expanding the exponentials and regrouping, as we did in deriving (11), we obtain without difficulty the formulas

$$\cosh z = \cosh x \cos y + i \sinh x \sin y \tag{17}$$

$$\sinh z = \sinh x \cos y + i \cosh x \sin y \tag{18}$$

In particular, by setting $x = 0$, we find

$$\cosh iy = \cos y \tag{19}$$

$$\sinh iy = i \sin y \tag{20}$$

The remaining hyperbolic functions are defined from $\cosh z$ and $\sinh z$ via the usual identities.

The logarithm of z we define implicitly as the function $w = \ln z$, which satisfies the equation

$$e^w = z \tag{21}$$

If we let $w = u + w$ and $z = re^{i\theta}$, Eq. (21) becomes

$$e^{u+iv} = e^{u}e^{iv} = re^{i\theta}$$

Hence, $e^{u} = r$ or $u = \ln r$ and $v = \theta$. Thus,

$$\begin{aligned} w = u + w &= \ln r + i\theta \\ &= \ln |z| + i \arg z \end{aligned} \tag{22}$$

If we let θ_1 be the **principal argument** of z, i.e., the particular argument of z which lies in the interval $-\pi < \theta \leq \pi$, Eq (22) can be written

$$\ln z = \ln |z| + i(\theta_1 + 2n\pi) \qquad n = 0, \pm 1, \pm 2, \tag{22a}$$

which shows that the logarithmic function is infinitely many-valued For any particular value of n, a unique branch of the function is determined, and the logarithm becomes effectively single-valued If $n = 0$, the resulting branch of the logarithmic function is called the **principal value.** Any particular branch of the logarithmic function is analytic, for, differentiating the definitive relation $z = e^{w}$, we have

$$\frac{dz}{dw} = e^{w} = z$$

$$\text{or} \qquad \frac{dw}{dz} \equiv \frac{d(\ln z)}{dz} = \frac{1}{z}$$

For a particular value of n the derivative of $\ln z$ thus exists for all $z \neq 0$

By means of (22a) the familiar laws of logarithms which hold for real variables can be established for complex variables as well. For example, to show that

$$\ln \frac{z_1}{z_2} = \ln z_1 - \ln z_2$$

$$\text{let} \qquad z_1 = r_1 e^{i\theta_1} \qquad \text{and} \qquad z_2 = r_2 e^{i\theta_2}$$

where θ_1 and θ_2 are the principal arguments of z_1 and z_2, respectively. Then,

$$\begin{aligned} \ln z_1 - \ln z_2 &= [\ln r_1 + i(\theta_1 + 2n_1\pi)] - [\ln r_2 + i(\theta_2 + 2n_2\pi)] \\ &= [\ln r_1 - \ln r_2] + i[(\theta_1 - \theta_2) + 2(n_1 - n_2)\pi] \\ &= \ln \frac{r_1}{r_2} + i[(\theta_1 - \theta_2) + 2n_3\pi] \\ &= \ln \left| \frac{z_1}{z_2} \right| + i \arg \frac{z_1}{z_2} \\ &= \ln \frac{z_1}{z_2} \end{aligned}$$

General powers of z are defined by the formula

$$z^{\alpha} = e^{\alpha \ln z} \tag{23}$$

which generalizes a familiar result for real variables which we frequently found useful in solving linear first-order differential equations Since $\ln z$ is infinitely many-valued, so, too, is z^{α}, in general Specifically,

$$z^\alpha = e^{\alpha \ln z} = e^{\alpha[\ln |z| + i(\theta_1 + 2n\pi)]} = e^{\alpha \ln |z|} e^{\alpha\theta_1 i} e^{2n\alpha\pi i}$$

The last factor in this product clearly involves infinitely many different values unless α is a rational number, say p/q; in which case, as we saw in our discussion of de Moivre's theorem in Sec 14.3, there are only q distinct values.*

EXAMPLE 3

What is the principal value of $(1 + i)^{2-i}$?

By definition,

$$\begin{aligned}(1 + i)^{2-i} &= e^{(2-i)\ln(1+i)} \\ &= e^{(2-i)[\ln\sqrt{2} + i(\pi/4 + 2n\pi)]}\end{aligned}$$

The principal value of this, obtained by taking $n = 0$, is

$$\begin{aligned}e^{(2-i)(\ln\sqrt{2} + i\pi/4)} &= e^{(2\ln\sqrt{2} + \pi/4) + i(-\ln\sqrt{2} + \pi/2)} \\ &= e^{\ln 2 + \pi/4}\left[\cos\left(\frac{\pi}{2} - \ln\sqrt{2}\right) + i\sin\left(\frac{\pi}{2} - \ln\sqrt{2}\right)\right] \\ &= e^{\ln 2 + \pi/4}[\sin(\ln\sqrt{2}) + i\cos(\ln\sqrt{2})] \\ &= e^{1\ 4785}(\sin 0\ 3466 + i\cos 0\ 3466) \\ &= 1\ 490 + 4\ 126i\end{aligned}$$

The inverse trigonometric and hyperbolic functions we define implicitly. For instance,

$$w = \cos^{-1} z$$

we define as the value or values of w which satisfy the equation

$$z = \cos w = \frac{e^{iw} + e^{-iw}}{2}$$

From this, by obvious steps, we obtain successively

$$e^{2iw} - 2ze^{iw} + 1 = 0$$

$$e^{iw} = z \pm \sqrt{z^2 - 1}$$

and, finally, by taking logarithms and solving for w,

$$w = \cos^{-1} z = -i\ln\left(z \pm \sqrt{z^2 - 1}\right) \tag{24}$$

Since the logarithm is infinitely many-valued, so, too, is $\cos^{-1} z$

Similarly, we can obtain the formulas

$$\sin^{-1} z = -i\ln\left(iz \pm \sqrt{1 - z^2}\right) \tag{25}$$

$$\tan^{-1} z = \frac{i}{2}\ln\frac{i + z}{i - z} \tag{26}$$

$$\cosh^{-1} z = \ln\left(z \pm \sqrt{z^2 - 1}\right) \tag{27}$$

$$\sinh^{-1} z = \ln\left(z \pm \sqrt{z^2 + 1}\right) \tag{28}$$

$$\tanh^{-1} z = \frac{1}{2}\ln\frac{1 + z}{1 - z} \tag{29}$$

---

* However, in the particular case $z = e$ the expression $z^\alpha = e^\alpha$ is single-valued for all values of α, rational or not, since $e^{\alpha_r + i\alpha_i}$ was defined simply as $e^{\alpha_r}(\cos\alpha_i + i\sin\alpha_i)$, which is clearly a unique complex number

From these, after their principal values have been suitably defined by choosing the positive square root and the principal value of the logarithm in each case, the usual differentiation formulas can be obtained without difficulty.

EXERCISES

1 Prove that $\cos^2 z + \sin^2 z = 1$.

2 Prove that $\cos(z_1 \pm z_2) = \cos z_1 \cos z_2 \mp \sin z_1 \sin z_2$

3 Prove that $\sin(z_1 \pm z_2) = \sin z_1 \cos z_2 \pm \cos z_1 \sin z_2$

4 Prove that $d(\cos z)/dz = -\sin z$

5 Prove that $d(\sin z)/dz = \cos z$

6 Express each of the following in the form $a + ib$, where a and b are decimal fractions:

a $\sin(2 - i)$ b $\cosh(1 + i)$ c $\sinh(2 + 3i)$

d The principal value of $\ln(-3 + 4i)$

e The principal value of $(1 - i)^{2-3i}$

7 Show that the various values of $(1 + i)^{1-i}$ differ only in their lengths

8 Prove that there is no value of z for which $e^z = 0$

9 If $g(x,y)$ is a real function of x and y, what is $|e^{ig(x,y)}|$?

10 Prove that $\overline{e^z} = e^{\bar{z}}$

11 Prove that $\overline{\cos z} = \cos \bar{z}$

12 Is $\overline{\ln z} = \ln \bar{z}$?

13 Is $\overline{\sin z} = \sin \bar{z}$?

14 Prove that the only zeros of $\cos z$ are the values $\pm\pi/2$, $\pm 3\pi/2$, $\pm 5\pi/2$,

15 Find all solutions of the equation $\sin z = 3$

16 Find all solutions of the equation $\cosh z = -2$

17 Find all solutions of the equation $e^z = -2$

18 By inspection, $e^0 > 0$ and $e^{i\pi} < 0$, yet by Exercise 8 there is no value of z for which $e^z = 0$ even though e^z is everywhere continuous Explain

19 Show that Rolle's theorem fails to hold for the function $e^{iz} - 1$, even though the conditions of the theorem appear to be satisfied with respect to the two values $z = 0$ and $z = 2\pi$. Explain

20 Show that $|\sin z|^2 = \sin^2 x + \sinh^2 y$ and that $|\cos z|^2 = \cos^2 x + \sinh^2 y$ What is $|\sinh z|^2$? What is $|\cosh z|^2$?

21 If $z = x + iy$, show that $|\sinh y| \leqq |\sin z| \leqq \cosh y$.

22 If $z = x + iy$, show that $|\sinh y| \leqq |\cos z| \leqq \cosh y$

23 a If $|z| \leqq 1$, show that $|\sin z| \leqq \frac{6}{5}|z|$

b Obtain an upper bound for $|\cos z|$, given that $|z| \leqq 1$

24 Show that $\sin \bar{z}$ and $\cos \bar{z}$ are not analytic functions of z

25 Prove that $\tan z = \dfrac{\sin 2x + i \sinh 2y}{\cos 2x + \cosh 2y}$

14.8

Integration in the complex plane

Line integrals in the complex plane are defined as follows: Let $f(z) = u(x,y) + iv(x,y)$ be any continuous function of z, analytic or not, and let C be a sectionally smooth arc joining the points A and B. Divide C into n subintervals Δs_k by the points z_k ($k = 1,$

2, . . . , $n - 1$), and let Δz_k be the infinitesimal chord determined by Δs_k Finally, in each subinterval choose an arbitrary point $\zeta_k = \xi_k + i\eta_k$ (Fig. 14.8). Then, if it exists, the limit of the sum

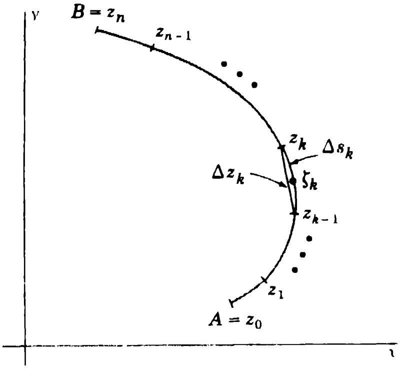


FIGURE 14.8 The subdivision of an arc preparatory to the definition of a line integral in the complex plane

(1)
$$\sum_{k=1}^{n} f(\zeta_k)\,\Delta z_k$$

as n becomes infinite in such a way that the length of each chord Δz_k approaches zero is called the **line integral** of $f(z)$ along C:

(2)
$$\int_C f(z)\,dz = \lim_{n\to\infty} \sum_{k=1}^{n} f(\zeta_k)\,\Delta z_k$$

In the special case when A and B coincide and C is a closed curve, the integral in (2) is often called a **contour integral** and is sometimes represented by the symbol

$$\oint f(z)\,dz$$

In working with complex line integrals it is frequently necessary to establish bounds on their absolute values To do this, let us return to the definitive sum (1) and apply to it the fundamental fact that the absolute value of a sum of complex numbers is less than or equal to the sum of their absolute values [Eq. (8), Sec 14 4]. Then,

$$\left|\sum_{k=1}^{n} f(\zeta_k)\,\Delta z_k\right| \leq \sum_{k=1}^{n} |f(\zeta_k)\,\Delta z_k| = \sum_{k=1}^{n} |f(\zeta_k)|\,|\Delta z_k|$$

the last equality following from the fact that the absolute value of a product is equal to the product of the absolute values [Eq (5), Sec. 14.4]. As $n \to \infty$, this yields a corresponding inequality for the integrals which are the limits of the respective sums.

(3)
$$\left|\int_C f(z)\,dz\right| \leq \int_C |f(z)|\,|dz|$$

The integral on the right is the real line integral

$$\int_C \sqrt{u^2 + v^2}\sqrt{(dx)^2 + (dy)^2} = \int_C \sqrt{u^2 + v^2}\, ds$$

where ds is the differential of arc length on C, which of course exists since C is assumed to be sectionally smooth. In particular, if $f(z) \equiv 1$, we have the simple but important result

$$\int_C |dz| = \int_C ds = L \tag{4}$$

where L is the length of the path of integration Since $f(z)$ is assumed to be continuous on the path of integration, including the end points A and B, it follows that $f(z)$ is a bounded function of z on the path of integration; that is, that there exists a constant M such that $|f(z)| \leqq M$ for all values of z on C Hence we have, from (3),

$$\left|\int_C f(z)\, dz\right| \leqq \int_C |f(z)|\, |dz| \leqq \int_C M|dz| = M \int_C |dz|$$

Therefore, using (4), we obtain the important inequality

$$\left|\int_C f(z)\, dz\right| \leqq ML \tag{5}$$

where M is any bound for $|f(z)|$ on C, and L is the length of the path of integration

Complex line integrals can readily be expressed in terms of real integrals For the sum (1) can be written

$$\sum_{k=1}^{n} [u(\xi_k,\eta_k) + iv(\xi_k,\eta_k)](\Delta x_k + i\, \Delta y_k) = \sum_{k=1}^{n} [u(\xi_k,\eta_k)\, \Delta x_k - v(\xi_k,\eta_k)\, \Delta y_k] + i \sum_{k=1}^{n} [v(\xi_k,\eta_k)\, \Delta x_k + u(\xi_k,\eta_k)\, \Delta y_k]$$

and, in the limit, the last expression yields the relation

$$\begin{aligned} \int_C f(z)\, dz &= \int_C u\, dx - v\, dy + i \int_C v\, dx + u\, dy \\ &= \int_C (u + iv)(dx + i\, dy) \end{aligned} \tag{6}$$

From (6) and the known properties of real line integrals (Sec. 12 4) or directly from the definition (2), it is easy to see that, when the same path of integration is used in each integral, we have

$$\int_A^B f(z)\, dz = -\int_B^A f(z)\, dz \tag{7}$$

$$\int_A^B kf(z)\, dz = k \int_A^B f(z)\, dz \tag{8}$$

$$\int_A^B [f(z) \pm g(z)]\, dz = \int_A^B f(z)\, dz \pm \int_A^B g(z)\, dz \tag{9}$$

and, if P is a third point on the arc AB,

$$\int_A^B f(z)\, dz = \int_A^P f(z)\, dz + \int_P^B f(z)\, dz \tag{10}$$

EXAMPLE 1

If C is a circle of radius r and center z_0, and if n is an integer, what is the value of

$$\int_C \frac{dz}{(z - z_0)^{n+1}}$$

For convenience, let us make the substitution $z - z_0 = re^{i\theta}$, noting that θ ranges from 0 t 2π as z ranges around the circle C (Fig 14.9) Then $dz = rie^{i\theta}\,d\theta$, and the integral becomes

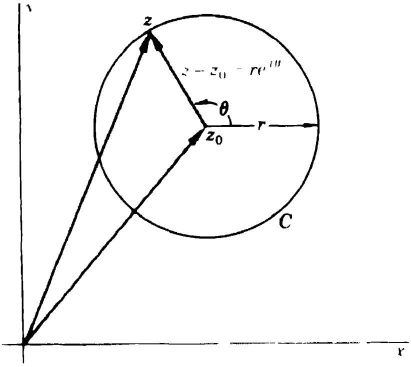


FIGURE 14.9 The circle $z - z_0 = re^{i\theta}$

$$\int_0^{2\pi} \frac{rie^{i\theta}\,d\theta}{r^{n+1}e^{i(n+1)\theta}} = \frac{i}{r^n}\int_0^{2\pi} e^{-in\theta}\,d\theta$$

If $n = 0$, this reduces to

$$i\int_0^{2\pi} = 2\pi i$$

On the other hand, if $n \neq 0$, we have

$$\frac{i}{r^n}\int_0^{2\pi} (\cos n\theta - i \sin n\theta)\,d\theta = 0$$

This is an important result to which we shall have occasion to refer from time to time

The form of the real line integrals in (6) suggests that Green's lemma (Theorem 1, Sec. 12.4) and the related results in Theorems 5 and 6, Sec. 12 5, may be useful in studying line integration in the complex plane, and this is indeed the case Hence, for ease of reference, we repeat this important material, appropriately specialized to the two-dimensional applications we now have in mind:*

THEOREM 1

If R is a region, either simply or multiply connected, whose boundary C is sectionally smooth and if $P(x,y)$, $Q(x,y)$, $\frac{\partial P}{\partial y}$, and $\frac{\partial Q}{\partial x}$ are continuous in and on the boundary of R, then

---

* To avoid confusion with u and v in the standard notation for a function of a complex variable, namely, $f(z) = u + iv$, we here use P and Q in place of the symbols U and V we used in Chap 12.

$$\int_C P\,dx + Q\,dy = \iint_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy$$

where the integration is taken around C in the positive direction with respect to the interior of R.

THEOREM 2

In any region where $\int P(x,y)\,dx + Q(x,y)\,dy$ is independent of the path, the partial derivatives of the function

$$\phi(x,y) = \int_{a,b}^{x,y} P(x,y)\,dx + Q(x,y)\,dy$$

are

$$\frac{\partial \phi}{\partial x} = P(x,y) \qquad \text{and} \qquad \frac{\partial \phi}{\partial y} = Q(x,y)$$

THEOREM 3

If $\dfrac{\partial Q}{\partial x} = \dfrac{\partial P}{\partial y}$ at all points of a simply connected region R, then in R the integral

$$\int P(x,y)\,dx + Q(x,y)\,dy$$

is independent of the path, and conversely

As a first application of Green's lemma, we have Cauchy's theorem, perhaps the most fundamental and far-reaching result in the theory of analytic functions

THEOREM 4

If R is a region, either simply or multiply connected, whose boundary C is sectionally smooth and if $f(z)$ is analytic and $f'(z)$ is continuous within and on the boundary of R, then

$$\int_C f(z)\,dz = 0$$

PROOF We begin by recalling from Eq (6) that

$$\int_C f(z)\,dz = \int_C u\,dx - v\,dy + i\int_C v\,dx + u\,dy$$

Now the hypothesis that $f'(z)$ is continuous means that the partial derivatives $\dfrac{\partial u}{\partial x}, \dfrac{\partial u}{\partial y}, \dfrac{\partial v}{\partial x}, \dfrac{\partial v}{\partial y}$ exist and are continuous throughout R Hence, Green's lemma can be applied to each of the line integrals on the right of the last expression, giving

$$\int_C f(z)\,dz = \iint_R \left(-\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) dx\,dy + i\iint_R \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) dx\,dy$$

However, u and v necessarily satisfy the Cauchy-Riemann equations, since, by hypothesis, $f(z)$ is analytic. Therefore, the integrand of each of the double integrals vanishes identically in R, leaving

$$\int_C f(z)\,dz = 0 \qquad \text{as asserted.}$$

The last theorem can be proved without making use of the hypothesis that $f'(z)$ is continuous * The French mathematician Edouard Goursat (1858–1936) was the first to do this, and in his honor the more general form of the result is usually referred to as the **Cauchy-Goursat theorem.**

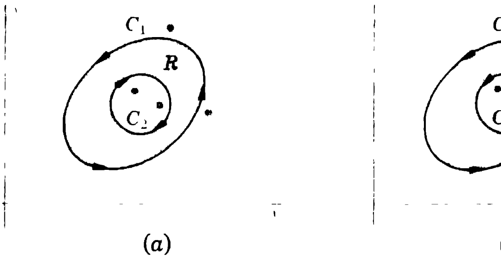


FIGURE 14 10 Contours which can be continuously deformed into each other

In particular, if $f(z)$ is analytic in and on the boundary of the region R between two simple closed curves, we have, from the Cauchy-Goursat theorem,

$$\int_{C_1} f(z)\,dz + \int_{C_2} f(z)\,dz = 0$$

provided that each curve is traversed in the positive direction, as shown in Fig 14 10a On the other hand, if we reverse the direction of integration around the inner curve C_2 and transpose the resultant integral, we obtain

$$\int_{C_1} f(z)\,dz = \int_{C_2} f(z)\,dz$$

each integration now being performed in the counterclockwise sense, as shown in Fig 14 10b Since there may be points in the interior of C_2 (which, of course, is not a part of R) where $f(z)$ is not analytic, we cannot assert that either of these integrals is zero However, we have shown that they both have the same value This result can be summarized in the highly important **principle of the deformation of contours:**

THEOREM 5

The line integral of an analytic function around any closed curve C_1 is equal to the line integral of the same function around any other closed curve C_2 into which C_1 can be continuously deformed without passing through a point where $f(z)$ is nonanalytic

If $f(z)$ is analytic throughout a simply connected region R, then, according to the Cauchy-Goursat theorem,

$$\int_C f(z)\,dz = 0$$

for every simple closed curve C in R. But, as we saw in the discussion which led to Theorem 6, Sec. 12 5, this implies that the line integral of $f(z)$ between any two points A and B in R is independent of the path. On the other hand, in multiply connected regions this observation is not necessarily true, since two different paths joining A and B might form a closed path encircling one of the inner boundaries of R and there is no assurance that the integral of $f(z)$ around such a path is zero Thus, summarizing, we have the following theorem:

---

* See, for example, E G Phillips, "Functions of a Complex Variable," pp 89–92, Interscience Publishers, Inc , New York, 1945

THEOREM 6

In any simply connected region where $f(z)$ is analytic, the integral $\int f(z)\,dz$ is independent of the path

Using Theorems 2 and 3 we can establish the following interesting result

THEOREM 7

If $u(x,y)$ is a solution of Laplace's equation in a region R, then in R there exists an analytic function having u as its real part, namely, $f(z) = u + iv$, where

$$v(x,y) = \int_{a,b}^{x,y} -\frac{\partial u}{\partial y}\,dx + \frac{\partial u}{\partial x}\,dy$$

and the path of integration from (a,b) to (x,y) lies entirely in R.

PROOF Suppose first that R is simply connected Then in R the integral defining v is independent of the path between the arbitrary fixed point (a,b) and the variable point (x,y), since the condition for independence provided by Theorem 3 is in this case

$$\frac{\partial\left(\dfrac{\partial u}{\partial x}\right)}{\partial x} = \frac{\partial\left(-\dfrac{\partial u}{\partial y}\right)}{\partial y} \qquad \text{or} \qquad \frac{\partial^2 u}{\partial x^2} = -\frac{\partial^2 u}{\partial y^2}$$

which is true because of the hypothesis that u satisfies Laplace's equation. Theorem 2 can, therefore, be applied to the integral which defines v, and we have

$$\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} \qquad \text{and} \qquad \frac{\partial v}{\partial y} = \frac{\partial u}{\partial x}$$

These are precisely the Cauchy-Riemann equations, which, if the derivatives are continuous, are the conditions that $f(z) = u + iv$ be an analytic function But $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$, and hence $\frac{\partial v}{\partial y}$ and $-\frac{\partial v}{\partial x}$, to which these are respectively equal, must be continuous, since the second partial derivatives $\frac{\partial^2 u}{\partial x^2}$ and $\frac{\partial^2 u}{\partial y^2}$ are known to exist. Hence, if R is simply connected, $f(z) = u + iv$ is analytic, as asserted

On the other hand, if R is multiply connected, then, by the principle of the deformation of contours, the possible values of v differ at most by constants independent of the end points And, clearly, a constant added to v will not affect the analyticity of $u + iv$ This completes the proof of the theorem.

One of the most important consequences of Cauchy's theorem is what is known as **Cauchy's integral formula:**

THEOREM 8

If $f(z)$ is analytic within and on the boundary C of a simply connected region R whose boundary C is sectionally smooth and if z_0 is any point in the interior of R, then

$$f(z_0) = \frac{1}{2\pi i}\int_C \frac{f(z)}{z - z_0}\,dz$$

the integration around C being taken in the positive sense.

PROOF Let C_0 be a circle with center at z_0 and radius ρ small enough so that C_0 lies entirely in R (Fig. 14.11). Now, by hypothesis, $f(z)$ is analytic everywhere within R. Hence, the function $f(z)/(z - z_0)$ is analytic everywhere within R except at the one point $z = z_0$. In particular, it is analytic everywhere in the region R' between C and C_0. Hence, by Theorem 5, the integral of this function around C is equal to its integral around C_0. That is,

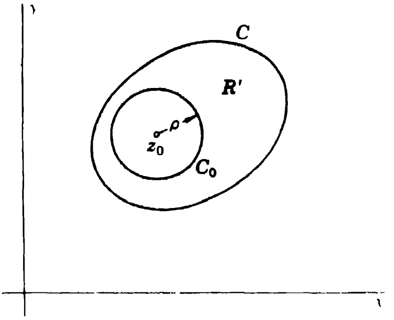


FIGURE 14.11 The circle C_0 used in the proof of Cauchy's integral formula

$$\int_C \frac{f(z)}{z - z_0}\,dz = \int_{C_0} \frac{f(z)}{z - z_0}\,dz = \int_{C_0} \frac{f(z_0) + [f(z) - f(z_0)]}{z - z_0}\,dz$$

$$(11) \qquad = f(z_0)\int_{C_0} \frac{dz}{z - z_0} + \int_{C_0} \frac{f(z) - f(z_0)}{z - z_0}\,dz$$

By Example 1, the first integral on the right is equal to $2\pi i$. Hence, the assertion of the theorem will be established if we can show that the last integral vanishes. To do this, we observe that

$$(12) \qquad \left|\int_{C_0} \frac{f(z) - f(z_0)}{z - z_0}\,dz\right| \leqq \int_{C_0} \frac{|f(z) - f(z_0)|}{|z - z_0|}\,|dz|$$

On C_0 we have $|z - z_0| = \rho$. Moreover, since $f(z)$ is analytic and hence continuous, it follows that, for any $\epsilon > 0$, there exists a δ such that

$$|f(z) - f(z_0)| < \epsilon \qquad \text{provided } |z - z_0| \equiv \rho < \delta$$

Choosing the radius ρ to be less than δ and inserting these estimates in the right member of (12), we therefore have

$$\left|\int_{C_0} \frac{f(z) - f(z_0)}{z - z_0}\,dz\right| < \int_{C_0} \frac{\epsilon}{\rho}\,|dz| = \frac{\epsilon}{\rho}\int_{C_0} |dz| = \frac{\epsilon}{\rho}\,2\pi\rho = 2\pi\epsilon$$

Since the integral on the left is independent of ϵ, yet cannot exceed $2\pi\epsilon$, which can be made arbitrarily small, it follows that the absolute value of th integral, and hence the integral itself, is zero. Thus, (11) reduces to

$$\int_{C_0} \frac{f(z)}{z - z_0}\,dz = f(z_0)2\pi i + 0$$

whence,

$$f(z_0) = \frac{1}{2\pi i}\int_{C_0} \frac{f(z)}{z - z_0}\,dz$$

as asserted Cauchy's integral formula is also true for multiply connected regions, but we shall leave as an exercise the easy modification of our proof required to establish this fact

EXAMPLE 2

Find the values of $\int_C \frac{e^z}{z^2 + 1}\,dz$ if C is a circle of unit radius with center at (a) $z = i$ and (b) $z = -i$.

In (a) we think of the integral as written in the form

$$\int_C \frac{e^z}{z + i} \cdot \frac{dz}{z - i}$$

and identify z_0 as i and $f(z)$ as $e^z/(z + i)$ The function $f(z)$ is analytic everywhere within and on the given circle of unit radius around $z = i$ (In fact, it is analytic everywhere except at $z = -i$) Therefore, we can apply Cauchy's integral formula, getting

$$\int_C \frac{e^z}{z + i}\,\frac{dz}{z - i} = 2\pi i f(i) = 2\pi i \frac{e^i}{2i} = \pi(\cos 1 + i \sin 1)$$

In (b) we identify z_0 as $-i$ and $f(z)$ as $e^z/(z - i)$ Then Cauchy's integral formula gives immediately

$$\int_C \frac{e^z}{z - i} \cdot \frac{dz}{z + i} = 2\pi i f(-i) = 2\pi i \frac{e^{-i}}{-2i} = -\pi(\cos 1 - i \sin 1)$$

From Cauchy's integral formula, which expresses the value of an analytic function at an interior point of a region R in terms of its values on the boundary of the region, we can readily obtain an expression for the derivative of a function at an interior point of R in terms of the boundary values of the function. In fact we have

$$\begin{aligned}
f'(z_0) &= \lim_{\Delta z_0 \to 0} \frac{f(z_0 + \Delta z_0) - f(z_0)}{\Delta z_0} \\
&= \lim_{\Delta z_0 \to 0} \frac{1}{\Delta z_0}\left[\frac{1}{2\pi i}\int_C \frac{f(z)\,dz}{z - (z_0 + \Delta z_0)} - \frac{1}{2\pi i}\int_C \frac{f(z)\,dz}{z - z_0}\right] \\
&= \lim_{\Delta z_0 \to 0} \frac{1}{\Delta z_0}\left[\frac{1}{2\pi i}\int_C f(z)\left(\frac{1}{z - (z_0 + \Delta z_0)} - \frac{1}{z - z_0}\right) dz\right] \\
&= \lim_{\Delta z_0 \to 0} \frac{1}{2\pi i}\int_C \frac{f(z)\,dz}{(z - z_0 - \Delta z_0)(z - z_0)}
\end{aligned}$$

Taking for granted that the limit of the integral is equal to the integral of the limit in the last expression and letting $\Delta z_0 \to 0$ in the integrand, we obtain the desired result.

$$f'(z_0) = \frac{1}{2\pi i}\int_C \frac{f(z)\,dz}{(z - z_0)^2}$$

That the limiting procedure is legitimate in this case can easily be established by showing that the absolute value of the difference

$$\frac{1}{2\pi i}\int_C \frac{f(z)\,dz}{(z - z_0 - \Delta z_0)(z - z_0)} - \frac{1}{2\pi i}\int_C \frac{f(z)\,dz}{(z - z_0)^2} \tag{13}$$

approaches zero as $\Delta z_0 \to 0$

Continuing in the same way, we obtain the additional formulas

$$f''(z_0) = \frac{2!}{2\pi i}\int_C \frac{f(z)\,dz}{(z - z_0)^3}$$

$$f'''(z_0) = \frac{3!}{2\pi i}\int_C \frac{f(z)\,dz}{(z - z_0)^4}$$

$$\cdot \qquad \cdot \quad \cdot \quad \cdot \quad \cdot$$

These results could all have been obtained formally by repeated differentiation of Cauchy's integral formula with respect to the parameter z_0

From the preceding discussion we conclude not only that an analytic function possesses derivatives of all orders but also that each derivative is itself analytic, since it, too, possesses a derivative. This completes the proof of the following theorem

THEOREM 9

If $f(z)$ is analytic throughout a closed, simply connected region R, then, at any interior point z_0 of R, the derivatives of $f(z)$ of all orders exist and are analytic Moreover,

$$f^{(n)}(z_0) = \frac{n!}{2\pi i}\int_C \frac{f(z)\,dz}{(z - z_0)^{n+1}}$$

where C is the boundary of R.

It is interesting to note that functions of a real variable do not in general possess the derivative properties described by Theorem 9, for, at particular points, a function of a real variable may possess one or more derivatives without the derivatives of all orders existing For instance, at the origin the function $x^{7/3}$ possesses a first and a second derivative but no derivatives of higher order.

Using Theorem 9, we can now prove the converse of Cauchy's theorem, which is known as **Morera's theorem**:*

THEOREM 10

If $f(z)$ is continuous in a region R and if $\int_C f(z)\,dz = 0$ for every simple closed curve C which can be drawn in R, then $f(z)$ is analytic in R

PROOF To prove this, we observe, as in the proof of Theorem 6, Sec 12 5, that, if the line integral of $f(z)$ around every closed curve in R is zero, then the line integral of $f(z)$ between a fixed point z_0 and a variable point z in R is independent of the path and, hence, is a function of z alone, say

---

* Named for the Italian mathematician Giacinto Morera (856–1909).

$$F(z) = \int_{z_0}^{z} f(z)\,dz$$

If we let $f(z) = u + iv$ and $F(z) = U + iV$, this can be written

$$F(z) = U + iV = \int_{x_0,y_0}^{x,y} u\,dx - v\,dy + i\int_{x_0,y_0}^{x,y} v\,dx + u\,dy$$

or, equating real and imaginary parts,

$$U = \int_{x_0,y_0}^{x,y} u\,dx - v\,dy \qquad \text{and} \qquad V = \int_{x_0,y_0}^{x,y} v\,dx + u\,dy$$

By Theorem 2, each of these integrals can be differentiated partially with respect to x and y, and we find

$$\frac{\partial U}{\partial x} = u \qquad \frac{\partial U}{\partial y} = -v \qquad \frac{\partial V}{\partial x} = v \qquad \frac{\partial V}{\partial y} = u$$

From these it is obvious that

$$\frac{\partial U}{\partial x} = \frac{\partial V}{\partial y} \qquad \text{and} \qquad \frac{\partial U}{\partial y} = -\frac{\partial V}{\partial x}$$

or, in other words, that U and V, satisfy the Cauchy-Riemann equations Moreover, since u and v are continuous, because of the hypothesis that $f(z) = u + iv$ is continuous, it follows that $\frac{\partial U}{\partial x}, \frac{\partial U}{\partial y}, \frac{\partial V}{\partial x}, \frac{\partial V}{\partial y}$ are continuous Hence, $F(z) = U + iV$ is an analytic function whose derivative, in fact, is

$$F'(z) = \frac{\partial U}{\partial x} + i\frac{\partial V}{\partial x} = u + iv = f(z)$$

Being the derivative of an analytic function, $f(z)$ is therefore analytic, by Theorem 9, as asserted

Beginning with the formula for $f^{(n)}(z_0)$ provided by Theorem 9, we can now establish what is known as **Cauchy's inequality**:

THEOREM 11

If $f(z)$ is analytic within and on a circle of radius r with center at z_0, then

$$|f^{(n)}(z_0)| \leqq \frac{n!M}{r^n}$$

where M is the maximum value of $|f(z)|$ on C

PROOF From Theorem 9, we have

$$\begin{aligned} |f^{(n)}(z_0)| &= \left|\frac{n!}{2\pi i}\int_C \frac{f(z)\,dz}{(z-z_0)^{n+1}}\right| \\ &\leqq \frac{n!}{2\pi}\int_C \frac{|f(z)|\,|dz|}{|z-z_0|^{n+1}} \\ &\leqq \frac{n!}{2\pi}\frac{M}{r^{n+1}}\int_C |dz| \\ &= \frac{n!}{2\pi}\frac{M}{r^{n+1}}2\pi r \\ &= \frac{n!M}{r^n} \end{aligned}$$

as asserted

For the special case $n = 0$, Cauchy's inequality becomes

$$|f(z_0)| \leq M$$

which shows that, on every circle around z_0, no matter how small, $|f(z)|$ has a maximum value M which is at least as great as $f(z_0)$ In other words, we have the following result, usually referred to as the **maximum modulus theorem**:

THEOREM 12

The absolute value of a function $f(z)$ cannot have a maximum at any point where the function is analytic.

EXERCISES

1 Evaluate $\int_0^{3+i} z^2\,dz$, (a) along the line $y = x/3$, (b) along the real axis to 3 and then vertically to $3 + i$, and (c) along the imaginary axis to i and then horizontally to $3 + i$

2 Evaluate $\int_0^{3+i} (\bar{z})^2\,dz$ along each of the paths used in Exercise 1

3 Evaluate $\int_0^{1+i} (x^2 + iy)\,dz$ along the paths $y = x$ and $y = x^2$

4 Obtain an upper bound for the absolute value of the integral $\int_0^{1+i} e^{-z^2}\,dz$, (a) along $y = x$, (b) along $y = x^2$, and (c) along the real axis to 1 and then vertically to $1 + i$

5 Obtain an upper bound for the absolute value of the integral $\frac{1}{2\pi i}\int \frac{e^{2z}}{z^2 + 1}\,dz$ taken around the circle $|z| = 3$ What is the value of this integral if the path of integration is the circle $|z| = \frac{1}{2}$?

6 What is the value of $\int_C \frac{3z^2 + 7z + 1}{z + 1}\,dz$, (a) if C is the circle $|z + 1| = 1$? (b) if C is the ellipse $x^2 + 2y^2 = 8$? (c) if C is the circle $|z + i| = 1$?

7 What is the value of $\int_C \frac{z + 4}{z^2 + 2z + 5}\,dz$ (a) if C is the circle $|z| = 1$? (b) if C is the circle $|z + 1 - i| = 2$? (c) if C is the circle $|z + 1 + i| = 2$?

8 What is the value of $\int \frac{e^z}{(z + 1)^2}\,dz$ around the circle $|z - 1| = 3$?

9 What is the value of $\int \frac{z + 1}{z^3 - 2z^2}\,dz$, (a) around the circle $|z| = 1$? (b) around the circle $|z - 2 - i| = 2$? and (c) around the circle $|z - 1 - 2i| = 2$?

10 Show that Cauchy's integral formula is valid in multiply connected regions

11 If $u(x,y)$ is harmonic, i e., satisfies Laplace's equation, within the closed region bounded by a circle C, prove that the maximum value of $u(x,y)$ in R always occurs on C and not in the interior of R [Hint Apply the maximum modulus theorem to the function $e^{f(z)}$, where $f(z)$ is the analytic function having $u(x,y)$ as its real part]

12 Using Theorem 9, show that

$$\frac{x^n}{n!} = \frac{1}{2\pi i}\int_C \frac{e^{xz}}{z^{n+1}}\,dz$$

where C is any simple closed curve encircling the origin

13 Observing that the result of Exercise 12 can be written

$$\left(\frac{x^n}{n!}\right)^2 = \frac{1}{2\pi i}\int_C \frac{x^n e^{xz}}{n!\,z^{n+1}}\,dz$$

prove that

$$\sum_{n=0}^{\infty}\left(\frac{x^n}{n!}\right)^2 = \frac{1}{2\pi}\int_C e^{2x\cos\theta}\,d\theta$$

14 Complete the proof of Theorem 9 by showing that the absolute value of the difference (13) approaches zero as Δz_0 approaches zero.

15 **a** Taking C to be the circle defined by $z = Re^{i\theta}$ and letting $z_0 = re^{i\phi}$ $(r < R)$, show that Cauchy's integral formula becomes

$$f(re^{i\phi}) = \frac{1}{2\pi i}\int_C \frac{f(Re^{i\theta})}{Re^{i\theta} - re^{i\phi}}\,d\theta$$

b Show also that

$$\frac{1}{2\pi i}\int_C \frac{f(Re^{i\theta})}{Re^{i\theta} - \dfrac{R^2}{r}e^{i\phi}}\,d\theta = 0$$

c Finally, by subtracting these two integrals and equating real parts in the resulting equation, obtain **Poisson's formula**:

$$u(r,\phi) = \frac{1}{2\pi}\int_0^{2\pi}\frac{(R^2 - r^2)u(R,\theta)}{R^2 - 2Rr\cos(\theta - \phi) + r^2}\,d\theta$$

CHAPTER FIFTEEN

Infinite Series in the Complex Plane

15.1

Series of complex terms

Most of the definitions and theorems relating to infinite series of real terms can be applied with little or no change to series whose terms are complex To restate these briefly, let

(1) $$f_1(z) + f_2(z) + f_3(z) + \cdots + f_n(z) + \cdots$$

be a series whose terms are functions of the complex variable z Then the **partial sums** of this series are defined as the finite sums

$$S_1(z) = f_1(z)$$

$$S_2(z) = f_1(z) + f_2(z)$$

$$\cdots \quad \cdots$$

$$S_n(z) = f_1(z) + f_2(z) + \cdots + f_n(z)$$

The series (1) is said to converge to the **sum** $S(z)$ in a region R provided that, for all values of z in R, the limit of the nth partial sum $S_n(z)$ as n becomes infinite is $S(z)$

According to the technical definition of a limit, this requires that, for any $\epsilon > 0$, there should exist an integer N, depending in general on ϵ and on the particular value of z under consideration, such that

$$|S(z) - S_n(z)| < \epsilon \qquad \text{for all } n > N$$

The difference $S(z) - S_n(z)$ is evidently just the remainder after n terms $R_n(z)$, thus, the definition of convergence requires that the limit of $|R_n(z)|$, as n becomes infinite, should be zero A series which has a sum, as just defined, is said to be **convergent,** and the set of all values of z for which it converges is called the **region of convergence** of the series. A series which is not convergent is said to be **divergent.** If the absolute values of the terms in (1) form a convergent series

$$|f_1(z)| + |f_2(z)| + |f_3(z)| + \cdots + |f_n(z)| + \cdots$$

then (1) is said to be **absolutely convergent.** If the series (1) converges but is not absolutely convergent, it is said to be **conditionally convergent.** Absolute convergence is an important property because it is a sufficient though not necessary condition for ordinary convergence Moreover, the terms of an absolutely convergent series can be rearranged in any manner whatsoever without affecting the sum of the series, whereas rearranging the terms of a conditionally convergent series may alter the sum of the series or even cause the series to diverge. From the definition of convergence it is easy to prove the following theorem.

THEOREM 1

A necessary and sufficient condition that the series of complex terms

$$f_1(z) + f_2(z) + f_3(z) + \qquad + f_n(z) + \cdots$$

converge is that the series of the real parts and the series of the imaginary parts of these terms each converge. Moreover, if

$$\sum_{n=1}^{\infty} \mathcal{R}(f_n) \qquad \text{and} \qquad \sum_{n=1}^{\infty} \mathcal{I}(f_n)$$

converge to the respective functions $R(z)$ and $I(z)$, then the given series converges to $R(z) + iI(z)$

Of all the tests for the convergence of infinite series, the most useful is probably the familiar **ratio test,** which applies to series whose terms are complex as well as to series whose terms are real

THEOREM 2

For the series

$$f_1(z) + f_2(z) + f_3(z) + \qquad + f_n(z) + \cdots$$

let

$$\lim_{n\to\infty} \left| \frac{f_{n+1}(z)}{f_n(z)} \right| = |r(z)|$$

Then the given series converges absolutely for those values of z for which $0 \leq |r(z)| < 1$ and diverges for those values of z for which $|r(z)| > 1$. The values of z for which $|r(z)| = 1$ form the boundary of the region of convergence of the series, and at these points the ratio test provides no information about the convergence or divergence of the series.

EXAMPLE 1

Find the region of convergence of the series

$$1 + \frac{1}{2^2}\left(\frac{z+1}{z-1}\right) + \frac{1}{3^2}\left(\frac{z+1}{z-1}\right)^2 + \frac{1}{4^2}\left(\frac{z+1}{z-1}\right)^3 + \cdots$$

Applying the ratio test, we find

$$\left| \frac{f_{n+1}(z)}{f_n(z)} \right| = \left| \frac{\dfrac{1}{(n+1)^2}\left(\dfrac{z+1}{z-1}\right)^n}{\dfrac{1}{n^2}\left(\dfrac{z+1}{z-1}\right)^{n-1}} \right| = \left| \frac{n^2}{(n+1)^2}\,\frac{z+1}{z-1} \right|$$

As n becomes infinite, this ratio approaches $\left|\dfrac{z+1}{z-1}\right|$. Hence, the values of z for which the series surely converges are those in the region defined by the inequality

$$\left|\frac{z+1}{z-1}\right| < 1$$

that is, by $$|z+1| < |z-1|$$

Now $|z+1|$ is just the distance from z to the point -1, and $|z-1|$ is just the distance from z to the point 1. Hence, z is restricted to be nearer to the point -1 than to the point 1. In other words, z must lie in the left half of the complex plane. The boundary cases for which the test fails are the values of z which are equidistant from -1 and 1, that is, the values of z on the imaginary axis. But, for these points, the related series of absolute values is the convergent real series

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots$$

Hence, for all values of z on the imaginary axis, the given series, being absolutely convergent, is convergent; therefore, these points also belong to the region of convergence.

The sum or difference of two convergent series can be found by term-by-term addition or subtraction of the series. If two series converge absolutely, their product can be found by multiplying the series together as though they were polynomials. To establish conditions under which series can legitimately be integrated or differentiated term by term, however, the concept of **uniform convergence** is required.

DEFINITION 1

A series of functions is said to converge uniformly to the function $S(z)$ in a region R, either open or closed, if corresponding to an arbitrary $\epsilon > 0$ there exists a positive integer N, depending on ϵ but not on z, such that for every value of z in R

$$|S(z) - S_n(z)| < \epsilon \qquad \text{for all } n > N$$

In other words, if a series converges uniformly in a region R, then, corresponding to any $\epsilon > 0$, there exists an integer N such that *everywhere* in R the sum of the series $S(z)$ can be approximated with an error less than ϵ by using *no more than* N terms of the series. It may well be that fewer than N terms will suffice at most of the points of the region, but *nowhere* will more than N be required. This is in sharp contrast to ordinary convergence, for, in the neighborhood of certain points in a region of ordinary convergence, it may be that no limit can be set on the number of terms required to secure a prescribed degree of accuracy.

EXAMPLE 2

Discuss the convergence of the series

$$z^2 + \frac{z^2}{1+z^2} + \frac{z^2}{(1+z^2)^2} + \frac{z^2}{(1+z^2)^3} + \cdots$$

in the 90° sector bounded by the right halves of the lines $y = \pm x$ (Fig. 15.1

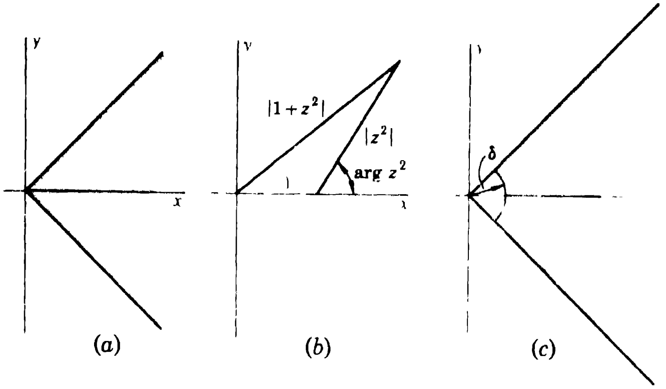


FIGURE 15.1 A 90° sector before and after modification to exclude its vertex

The given series is a geometric progression which will converge for all values of z for which the absolute value of the common ratio, i.e.,

$$|r| = \frac{1}{|1 + z^2|}$$

is less than 1. Now the angle of z is restricted, by hypothesis, to be between $-\pi/4$ and $\pi/4$, hence, the angle of z^2 must be between $-\pi/2$ and $\pi/2$. Therefore (Fig. 15.1b), for every z in the given region R, we have

$$|1 + z^2| \geqq 1 \qquad \text{and} \qquad \frac{1}{|1 + z^2|} \leqq 1$$

and the equality signs hold only for the value $z = 0$. Thus, the given series converges for all values of z in R, and its sum is

$$S(z) = \begin{cases} \dfrac{a}{1 - r} = \dfrac{z^2}{1 - 1/(1 + z^2)} = 1 + z^2 & z \neq 0 \\ 0 + 0 + 0 + 0 + \cdots = 0 & z = 0 \end{cases}$$

Now let an arbitrary $\epsilon > 0$ be given, and let us attempt to determine how many terms of the series must be taken in order that

$$|S(z) - S_n(z)| < \epsilon$$

This difference, i.e., the remainder after n terms of the series, is just the geometric progression

$$\frac{z^2}{(1 + z^2)^n} + \frac{z^2}{(1 + z^2)^{n+1}} + \frac{z^2}{(1 + z^2)^{n+2}} + \cdots$$

whose sum is

$$R_n(z) = \begin{cases} \dfrac{1}{(1 + z^2)^{n-1}} & z \neq 0 \\ 0 & z = 0 \end{cases}$$

Hence, our task is to find, if possible, a value of N such that

(2) $$|R_n(z)| = \frac{1}{|1 + z^2|^{n-1}} < \epsilon \qquad \text{for all } n > N \text{ and all } z \text{ in } R$$

Now $|1 + z^2| \leqq 1 + |z^2| = 1 + |z|^2$. Hence, overestimating the denominator of $|R_n(z)|$, we have

$$|R_n(z)| = \frac{1}{|1 + z^2|^{n-1}} \geqq \frac{1}{(1 + |z|^2)^{n-1}}$$

From this inequality we observe that if it should be impossible to find an integer N such that

(3) $$\frac{1}{(1 + |z|^2)^{n-1}} < \epsilon \qquad \text{for all } n > N \text{ and all } z \text{ in } R$$

then surely it will be impossible to find an integer N which will suffice to keep

$$|R_n(z)| < \epsilon$$

everywhere in R And this is indeed the case, for if we attempt to solve the inequality (3) for n, by obvious steps we find

$$(1 + |z|^2)^{n-1} > \frac{1}{\epsilon}$$

$$(n - 1) \ln (1 + |z|^2) > \ln \frac{1}{\epsilon} = -\ln \epsilon$$

$$n > 1 - \frac{\ln \epsilon}{\ln (1 + |z|^2)}$$

But, for values of z within the sector of the problem and sufficiently close to the origin, $\ln (1 + |z|^2)$ can be made arbitrarily close to $\ln 1$, that is, zero Hence, n is unbounded, and there exists no integer N for which (3) holds Since $|R_n(z)|$ is larger than the fraction in (3), it is clear that the fundamental requirement of uniform convergence (2) cannot be fulfilled Hence, the convergence of the given series in the original region is nonuniform

On the other hand, if we restrict z to the infinite region R', bounded by the given rays and a circular arc of small but fixed radius δ, as shown in Fig 15 1c, the series converges uniformly In fact, the law of cosines applied to Fig 15 1b gives

$$|1 + z^2|^2 = 1 + |z^2|^2 + 2|z^2| \cos (\arg z^2)$$

or, reducing the right-hand side by dropping the last term, which is surely nonnegative, since $-\pi/2 \leqq \arg z^2 \leqq \pi/2$,

$$|1 + z^2|^2 \geqq 1 + |z^2|^2 = 1 + |z^4|$$

Hence, underestimating the denominator of $|R_n(z)|$, we can write

$$|R_n(z)| = \frac{1}{|1 + z^2|^{n-1}} \leqq \frac{1}{(1 + |z|^4)^{(n-1)/2}}$$

From this it is clear that, if we can find an integer N such that

(4) $$\frac{1}{(1 + |z|^4)^{(n-1)/2}} < \epsilon \qquad \text{for all } n > N \text{ and all } z \text{ in } R'$$

then surely for the same N we shall have

(5) $$|R_n(z)| < \epsilon \qquad \text{for all } n > N \text{ and all } z \text{ in } R'$$

Hence, we attempt to solve the inequality in (4) for n

$$(1 + |z|^4)^{(n-1)/2} > \frac{1}{\epsilon}$$

$$\frac{n-1}{2} \ln (1 + |z|^4) > \ln \frac{1}{\epsilon} = -\ln \epsilon$$

$$n > 1 - \frac{2 \ln \epsilon}{\ln (1 + |z|^4)}$$

The most unfavorable case, i e , the largest possible value of the fraction on the right, occurs when $|z|$ is as small as possible But, in the modified region we are now considering, the smallest possible value of $|z|$ is δ, which yields

$$n > 1 - \frac{2 \ln \epsilon}{\ln (1 + \delta^4)}$$

If we choose N to be the first integer equal to or greater than the expression on the right, then (4) will surely hold But, as we observed above, if (4) is satisfied, so too is (5), and, hence, in the modified region R' the given series converges uniformly

Usually uniform convergence is established not by a direct application of the definition, as in Example 2, but by the so-called **Weierstrass M test:***

THEOREM 3

If a sequence of positive constants $\{M_n\}$ exists such that $|f_n(z)| \leqq M_n$ for all positive integers n and for all values of z in a given region R and if the series

$$M_1 + M_2 + M_3 + \qquad + M_n +$$

is convergent, then the series

$$f_1(z) + f_2(z) + f_3(z) + \qquad + f_n(z) +$$

converges uniformly in R

PROOF To prove this, we must show that for any $\epsilon > 0$ there exists an integer N independent of z, such that for all values of z in R the absolute value of the remainder after n terms in the series of the f's is less than ϵ whenever n exceeds N To do this, we note that

$$\begin{aligned} |R_n(z)| &= |f_{n+1}(z) + f_{n+2}(z) + \qquad | \\ &\leqq |f_{n+1}(z)| + |f_{n+2}(z)| + \\ &\leqq M_{n+1} + M_{n+2} + \end{aligned} \tag{6}$$

The last expression is just the remainder after n terms of the series of the M's. Since this series is convergent, by hypothesis, it follows that, for every $\epsilon > 0$, there exists an N such that this remainder is less than ϵ for all $n > N$ This value of N, arising as it does from a series of constants, is obviously independent of z. Moreover, from the inequality (6) it is clear that whenever n exceeds this N, $|R_n(z)| < \epsilon$ for all values of z in R Hence, the series of the f's is uniformly convergent, as asserted Incidentally, this theorem implies a comparison test which proves that the series of the f's is also absolutely convergent

The Weierstrass M test is merely a sufficient test, that is, there exist uniformly convergent series whose terms cannot be dominated by the respective terms of any convergent series of positive constants † The M test suffices for almost all applications, however

One useful property of uniformly convergent series is contained in the following theorem

THEOREM 4

If the terms of a uniformly convergent series are multiplied by any bounded function of z, the resulting series will also converge uniformly.

---

* Karl Weierstrass (1815–1897), a German mathematician, is often called the "father of modern rigor"

† One example of such a series will be found in Exercise 6

PROOF Let R be the region of uniform convergence of the series

$$f_1(z) + f_2(z) + f_3(z) + \cdots + f_n(z) + \cdots$$

and suppose that throughout R we have

$$|g(z)| \leqq M$$

Now, since the series of the f's converges uniformly, it follows that, corresponding to the infinitesimal ϵ/M, there exists an integer N such that

$$|f_{n+1}(z) + f_{n+2}(z) + \cdots| < \frac{\epsilon}{M} \qquad \text{for all } n > N \text{ and all } z \text{ in } R$$

Hence,

$$\begin{aligned} |g(z)f_{n+1}(z) + g(z)f_{n+2}(z) + \cdots| &= |g(z)|\,|f_{n+1}(z) + f_{n+2}(z) + \cdots| \\ &\leqq M|f_{n+1}(z) + f_{n+2}(z) + \cdots| \\ &\leqq M\frac{\epsilon}{M} \\ &= \epsilon \qquad \text{for all } n > N \text{ and all } z \text{ in } R \end{aligned}$$

But this is precisely the condition that the product series

$$g(z)f_1(z) + g(z)f_2(z) + g(z)f_3(z) + \cdots + g(z)f_n(z) + \cdots$$

be uniformly convergent

One important consequence of uniform convergence is embodied in the following theorem

THEOREM 5

The sum of a uniformly convergent series of continuous functions is a continuous function

PROOF Let

$$f(z) = f_1(z) + f_2(z) + f_3(z) + \cdots + f_n(z) + \cdots = S_n(z) + R_n(z)$$

be a uniformly convergent series in which each term is a continuous function of z, and let $\epsilon/3$ be an arbitrary infinitesimal Then, since the series converges uniformly, an integer N exists such that

$$|R_n(z)| < \frac{\epsilon}{3} \qquad \text{for all } n > N$$

and for all values of z in the region of uniform convergence In particular, if $\Delta_1 z$ is any increment such that $z + \Delta_1 z$ is still in the region of uniform convergence, we also have

$$|R_n(z + \Delta_1 z)| < \frac{\epsilon}{3} \qquad \text{for all } n > N$$

Moreover, since each term of the given series is a continuous function and since any *finite* sum of continuous functions is necessarily continuous, it follows that $S_n(z)$ is continuous and, hence, that there exists an increment $\Delta_2 z$ such that

$$|S_n(z + \Delta z) - S_n(z)| < \frac{\epsilon}{3} \qquad \text{for all } \Delta z\text{'s for which } |\Delta z| < |\Delta_2 z|$$

Now

$$
\begin{aligned}
|f(z+\Delta z)-f(z)| &= |[S_n(z+\Delta z)+R_n(z+\Delta z)]-[S_n(z)+R_n(z)]| \\
&\leqq |S_n(z+\Delta z)-S_n(z)|+|R_n(z+\Delta z)|+|R_n(z)|
\end{aligned}
$$

Hence, for all Δz's whose absolute values are less than the smaller of the quantities $|\Delta_1 z|$ and $|\Delta_2 z|$, it follows that

$$|f(z+\Delta z)-f(z)| < \frac{\epsilon}{3}+\frac{\epsilon}{3}+\frac{\epsilon}{3}=\epsilon$$

which is precisely what we mean by saying that $f(z)$ is continuous.

Theorem 5 makes no assertion about the sum of a series of continuous functions if the convergence is nonuniform However, specific examples make it clear that in such cases the sum need not be continuous For instance, Example 2, in which we found the sum of the series

$$z^2+\frac{z^2}{1+z^2}+\frac{z^2}{(1+z^2)^2}+\frac{z^2}{(1+z^2)^3}+\cdots$$

to be $f(z)=\begin{cases}1+z^2 & z\neq 0\\ 0 & z=0\end{cases}$

shows that the limit of a sum of continuous functions may be discontinuous if the convergence is nonuniform In fact, in the neighborhood of $z=0$, where the convergence is nonuniform, the sum jumps abruptly from $1+z^2$ to 0, even though every term of the series is a continuous function of z for all values of z except $z=\pm i$.

One of the most important properties of uniformly convergent series is given by the following theorem

THEOREM 6

The integral of the sum of a uniformly convergent series of continuous functions along any curve C lying entirely in the region of uniform convergence can be found by term-by-term integration of the series. Moreover, if each term of the series is analytic, so, too, is the sum

PROOF Let the given series be

$$f(z)=f_1(z)+f_2(z)+\cdots+f_n(z)+\cdots$$

Then, to establish the theorem we must show that

$$\int_C f(z)\,dz=\int_C f_1(z)\,dz+\int_C f_2(z)\,dz+\cdots+\int_C f_n(z)\,dz+\cdots$$

which, in accordance with the usual definition of convergence, requires that we prove the existence, for every $\epsilon>0$, of an integer N such that

$$\left|\int_C f(z)\,dz-\sum_{i=1}^{n}\int_C f_i(z)\,dz\right|<\epsilon \qquad \text{for all } n>N$$

Now for any *finite* sum it is true that the integral of a sum is equal to the sum of the integrals. Hence, the left member of the last inequality can be written

$$\left| \int_C f(z)\,dz - \int_C \sum_{i=1}^{n} f_i(z)\,dz \right| = \left| \int_C \left[f(z) - \sum_{i=1}^{n} f_i(z) \right] dz \right| = \left| \int_C R_n(z)\,dz \right|$$

Let L be the length of the path of integration. Then, from the uniform convergence of the given series, we know that there exists an integer N such that

$$|R_n(z)| < \frac{\epsilon}{L} \qquad \text{for all } n > N$$

and for all z's in the region of uniform convergence, in particular for all values of z on the path of integration C. If $n > N$, we can therefore write

$$\left| \int_C f(z)\,dz - \sum_{i=1}^{n} \int_C f_i(z)\,dz \right| = \left| \int_C R_n(z)\,dz \right| \leqq \int_C |R_n(z)|\,|dz|$$

$$< \frac{\epsilon}{L} \int_C |dz| = \frac{\epsilon}{L} L = \epsilon$$

which establishes the first part of the theorem

To establish the second part, we suppose that the region of uniform convergence R is either simply connected or has been made simply connected by suitable cross cuts Then, if each term f_i is analytic in R, it follows from Cauchy's theorem that the integral of each term around any simple closed curve in R (or its simply connected modification) is zero Hence, the integral of the sum $f(z)$ around any closed curve is zero, and, thus, by Morera's theorem, $f(z)$ is analytic This completes the proof of the theorem.

The companion result on the term-by-term differentiation of series is contained in the following theorem·

THEOREM 7

If $f(z)$ is the sum of a uniformly convergent series of analytic functions, then the derivative of $f(z)$ at any interior point of the region of uniform convergence can be found by term-by-term differentiation of the series

PROOF Let z be a general point of the region of uniform convergence R, and let C be a simple closed curve drawn around z in R. If we write the given series as

$$f(t) = f_1(t) + f_2(t) + \cdot \quad \cdot + f_n(t) + \cdot \cdot \cdot$$

where t stands for any of the values of z on C, we can multiply by the bounded function

$$\frac{1}{2\pi i(t-z)^2}$$

and, by Theorem 4, the resulting series

$$\frac{f(t)}{2\pi i(t-z)^2} = \frac{f_1(t)}{2\pi i(t-z)^2} + \frac{f_2(t)}{2\pi i(t-z)^2} + \cdot \cdot \cdot + \frac{f_n(t)}{2\pi (t-z)^2} +$$

will also converge uniformly. By Theorem 6, it can, therefore, be integrated term by term around C, giving

$$\frac{1}{2\pi i}\int_C \frac{f(t)\,dt}{(t-z)^2} = \frac{1}{2\pi i}\int_C \frac{f_1(t)\,dt}{(t-z)^2} + \frac{1}{2\pi i}\int_C \frac{f_2(t)\,dt}{(t-z)^2} + \cdots + \frac{1}{2\pi i}\int_C \frac{f_n(t)\,dt}{(t-z)^2} + \cdots$$

But these integrals, by the first generalization of Cauchy's formula (Theorem 9, Sec. 14.8), are precisely the derivatives of the respective terms of the given series at the point z. Hence,

$$f'(z) = f_1'(z) + f_2'(z) + \cdots + f_n'(z) + \cdots$$

which establishes the theorem.

It is interesting and important to note that Theorem 7 does not apply to series of functions of the real variable x. To justify term-by-term differentiation of such series, we require not uniform convergence of the original series, but rather uniform convergence of the series resulting from the term-by-term differentiation. More precisely, we have the following theorem, which is proved in most texts on advanced calculus.*

THEOREM 8

If $$f(x) = f_1(x) + f_2(x) + f_3(x) + \cdots + f_n(x) + \cdots$$

is a convergent series of functions of the real variable x, each of which possesses a continuous first derivative, then $f'(x)$ can be found by term-by-term differentiation, provided the series of the derivatives is uniformly convergent.

EXERCISES

1 Find the region of convergence of the series

$$1 + (z-i) + (z-i)^2 + (z-i)^3 + \cdots$$

2 Find the region of convergence of the series

$$\frac{1}{2(z+i)} + \frac{1}{2^2(z+i)^2} + \frac{1}{2^3(z+i)^3} + \frac{1}{2^4(z+i)^4} + \cdots$$

3 Find the region of convergence of the series

$$1 + \frac{1}{2^2}\left(\frac{\mathcal{R}(z)}{z+1}\right) + \frac{1}{3^2}\left(\frac{\mathcal{R}(z)}{z+1}\right)^2 + \frac{1}{4^2}\left(\frac{\mathcal{R}(z)}{z+1}\right)^3 + \cdots$$

4 Show that the entire region of convergence of the series of Example 2 consists of the exterior of the lemniscate $(x^2 - y^2 + 1)^2 + 4x^2y^2 = 1$ together with the origin.

5 Show that the series $x + x(1-x) + x(1-x)^2 + x(1-x)^3 + \cdots$ converges for $0 \leqq x < 2$, but that the convergence is nonuniform in any subinterval which contains the origin.

---

* See, for instance, A. E. Taylor, "Advanced Calculus," p. 602, Ginn and Company, Boston, 1955.

6 Show that the series

$$\frac{1}{1+x^2} - \frac{1}{2+x^2} + \frac{1}{3+x^2} - \frac{1}{4+x^2} + \cdots$$

converges uniformly over any interval of the x-axis, but that this cannot be established by the Weierstrass M test

7 Show that the series

$$\frac{z}{(0\cdot z+1)(z+1)} + \frac{z}{(z+1)(2z+1)} + \frac{z}{(2z+1)(3z+1)} + \frac{z}{(3z+1)(4z+1)} + \cdots$$

converges to 0 if $z = 0$ and to 1 if $z \neq 0$ Show that the convergence is nonuniform in the neighborhood of the origin, but uniform in the exterior of any circle with center at the origin

8 What is the region of convergence of the series $\sum_{n=1}^{\infty} \frac{e^{inz}}{n^{3/2}}$? Where does the series converge uniformly? Show that the series $\sum_{n=1}^{\infty} \frac{e^{inx}}{n^{3/2}}$ converges uniformly over any interval of the x-axis, but that it cannot be differentiated term by term for any value of x Explain

9 Can the sum of a nonuniformly convergent series of continuous functions be continuous?

10 Prove Theorem 1.

15.2

Taylor's expansion

Very often the series with which one has to deal in applications are those which are studied formally in elementary calculus under the name of *Taylor's series* Their systematic study begins with **Taylor's theorem**:*

THEOREM 1

If $f(z)$ is analytic throughout the region bounded by a simple closed curve C and if z and a are both interior to C, then

$$f(z) = f(a) + f'(a)(z-a) + f''(a)\frac{(z-a)^2}{2!} + \cdots + f^{(n-1)}(a)\frac{(z-a)^{n-1}}{(n-1)!} + R_n$$

where $$R_n = \frac{(z-a)^n}{2\pi i}\int_C \frac{f(t)\,dt}{(t-a)^n(t-z)}$$

PROOF We first note that Cauchy's integral formula can be written

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(t)\,dt}{t-z} = \frac{1}{2\pi i}\int_C \frac{f(t)}{t-a}\,\frac{1}{1-(z-a)/(t-a)}\,dt$$

Then, from this, applying the identity

$$\frac{1}{1-u} = 1 + u + u^2 + \cdots + u^{n-1} + \frac{u^n}{1-u}$$

---

* Named for the English mathematician Brook Taylor (1685–1731).

to the factor $\dfrac{1}{1-(z-a)/(t-a)}$

in the last integral, we have

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(t)}{t-a}\left[1 + \left(\frac{z-a}{t-a}\right) + \left(\frac{z-a}{t-a}\right)^2 + \cdots \right.$$
$$\left. + \left(\frac{z-a}{t-a}\right)^{n-1} + \frac{(z-a)^n/(t-a)^n}{1-(z-a)/(t-a)}\right] dt$$
$$= \frac{1}{2\pi i}\int_C \frac{f(t)\,dt}{t-a} + \frac{z-a}{2\pi i}\int_C \frac{f(t)\,dt}{(t-a)^2} + \cdots$$
$$+ \frac{(z-a)^{n-1}}{2\pi i}\int_C \frac{f(t)\,dt}{(t-a)^n} + \frac{(z-a)^n}{2\pi i}\int_C \frac{f(t)\,dt}{(t-a)^n(t-z)}$$

From the generalizations (Theorem 9, Sec 14 8) of Cauchy's integral formula it is evident that, except for the necessary factorials, the first n integrals in the last expression are precisely the corresponding derivatives of $f(z)$ evaluated at the point $z = a$. Hence,

$$f(z) = f(a) + f'(a)(z-a) + \cdots + f^{(n-1)}(a)\frac{(z-a)^{n-1}}{(n-1)!}$$
$$+ \frac{(z-a)^n}{2\pi i}\int_C \frac{f(t)\,dt}{(t-a)^n(t-z)}$$

which establishes the theorem.

By **Taylor's series** we mean the infinite expansion suggested by the last theorem, namely,

$$f(z) \sim f(a) + f'(a)(z-a) + f''(a)\frac{(z-a)^2}{2!} + \cdots$$
$$+ f^{(n-1)}(a)\frac{(z-a)^{n-1}}{(n-1)!} + \cdots$$

To show that this series actually converges to $f(z)$, we must show, as usual, that the absolute value of the difference between $f(z)$ and the sum of the first n terms of the series approaches zero as n becomes infinite From Taylor's theorem it is evident that this difference is

$$R_n(z) = \frac{(z-a)^n}{2\pi i}\int_C \frac{f(t)\,dt}{(t-a)^n(t-z)}$$

Accordingly, we must determine the values of z for which the absolute value of this integral approaches zero as n becomes infinite

To do this, let C_1 and C_2 be two circles of radii r_1 and r_2 having their centers at the point a and lying entirely in the interior of C (Fig 15.2). Since $f(z)$ is analytic throughout the interior of C, the entire integrand of $R_n(z)$ is analytic in the region between C and C_2, provided that z, like a, lies in the interior of C_2. Under these conditions, the integral around C can be replaced by the integral around C_2 If, in addition, z is

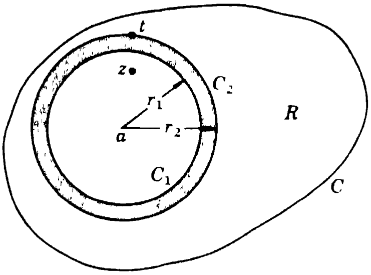


FIGURE 15.2
The circles C_1 and C_2 used in the proof of the convergence of Taylor's series

interior to C_1, then for all values of t on C_2 (the t's which are now involved in the integration) we have

$$|t - a| = r_2$$

$$|z - a| < r_1$$

$$|t - z| > r_2 - r_1$$

and $$|f(t)| \leqq M$$

where M is the maximum of $|f(z)|$ on C_2 Hence, overestimating factors in the numerator and underestimating factors in the denominator, we have

$$\begin{aligned}
|R_n(z)| &= \left| \frac{(z-a)^n}{2\pi i} \int_C \frac{f(t)\,dt}{(t-a)^n(t-z)} \right| \\
&\leq \frac{|z-a|^n}{|2\pi i|} \int_C \frac{|f(t)|\,|dt|}{|t-a|^n|t-z|} \\
&< \frac{r_1^n}{2\pi} \int_C \frac{M|dt|}{r_2^n(r_2-r_1)} \\
&= \frac{r_1^n M}{2\pi r_2^n(r_2-r_1)}\, 2\pi r_2 \\
&= M\left(\frac{r_1}{r_2}\right)^n \frac{r_2}{r_2-r_1}
\end{aligned}$$

Since $0 < r_1 < r_2$, the fraction $(r_1/r_2)^n$ approaches zero as n becomes infinite; therefore, the limit of $R_n(z)$ is zero Thus we have established the following important theorem

THEOREM 2

Taylor's series,

$$f(z) = f(a) + f'(a)(z-a) + f''(a)\frac{(z-a)^2}{2!} + f'''(a)\frac{(z-a)^3}{3!} + \cdots$$

is a valid representation of $f(z)$ at all points in the interior of any circle having its center at a and within which $f(z)$ is analytic.

The largest circle which can be drawn around $z = a$ such that $f(z)$ is analytic everywhere in its interior is called **the circle of convergence** of the Taylor's series of $f(z)$ about the point $z = a$.

The radius of this circle is called the **radius of convergence** of the series. Of course, this entire discussion applies without change to the case $a = 0$, which is usually called **Maclaurin's series.***

The preceding discussion established a circular region around the point $z = a$ within which the Taylor's series of $f(z)$ converges to $f(z)$. However, it did not provide any information about the behavior of the series outside the circle of convergence. Actually, the Taylor's series of $f(z)$ converges only within and possibly on the circle of convergence, and diverges everywhere outside this circle, as the following two theorems make clear

THEOREM 3

If the power series

$$a_0 + a_1(z - a) + a_2(z - a)^2 + a_3(z - a)^3 + \cdots$$

converges for $z = z_1$, it converges absolutely for all values of z such that $|z - a| < |z_1 - a|$ and uniformly for all values of z such that $|z - a| \leqq r < |z_1 - a|$. Moreover, the sum to which it converges is analytic

PROOF Since the given series converges when $z = z_1$, it follows that the terms of the series are bounded for this value of z That is, there exists a positive constant M such that

$$|a_n(z_1 - a)^n| = |a_n|\,|z_1 - a|^n \leqq M \qquad \text{for } n = 0, 1, 2, \ldots$$

Now let z_0 be any value of z such that

$$|z_0 - a| < |z_1 - a|$$

that is, let z_0 be any point nearer to a than z_1 is Then, for the general term of the series when $z = z_0$, we have

$$|a_n(z_0 - a)^n| = |a_n|\,|z_0 - a|^n = |a_n|\,|z_1 - a|^n \left|\frac{z_0 - a}{z_1 - a}\right|^n \leqq M \left|\frac{z_0 - a}{z_1 - a}\right|^n$$

If we set

$$\left|\frac{z_0 - a}{z_1 - a}\right| = k \tag{1}$$

where k is obviously less than 1, this shows that the absolute values of the terms of the series

$$a_0 + a_1(z_0 - a) + a_2(z_0 - a)^2 + a_3(z_0 - a)^3 + \cdots \tag{2}$$

are dominated, respectively, by the terms of the series of positive constants

$$M + Mk + Mk^2 + Mk^3 + \cdots \tag{3}$$

This is a geometric series whose common ratio k is numerically less than 1. It therefore converges and, hence, provides a comparison test which establishes the absolute convergence of the given series (2)

Unfortunately, the series (3) does not provide a test series which can be used in applying the Weierstrass M test to the series (2), because it is clear from (1) that the terms of the series (3) depend on z_0 However, for values of z_0 such that

---

* Named for the Scottish mathematician Colin Maclaurin (1698 1746), although another Scottish mathematician, James Stirling (1692–1770), anticipated by 25 years Maclaurin's use of this result

(4) $$|z_0 - a| \leqq r < |z_1 - a|$$

we have $$k = \left|\frac{z_0 - a}{z_1 - a}\right| \leqq \frac{r}{|z_1 - a|} = \lambda$$

and λ is clearly a positive constant less than 1 which *is* independent of z_0. Hence, for all values of z_0 satisfying the condition (4), the series (1) is dominated term by term by the convergent geometric series of positive constants

$$M + M\lambda + M\lambda^2 + M\lambda^3 + \cdots$$

and, therefore, by Theorem 3, Sec 15.1, the series (2) is uniformly convergent

Finally, since each term $a_n(z - a)^n$ is an analytic function and since any point in the interior of the circle $|z - a| = |z_1 - a|$ can be included within a circle of the form $|z - a| = r < |z_1 - a|$, it follows from the second part of Theorem 6, Sec 15 1, that, within the circle $|z - a| = |z_1 - a|$, the function to which the series converges is analytic.

Now, let α be the singular point of $f(z)$ nearest to the center of the expansion $z = a$, and suppose that the Taylor's series for $f(z)$ converges for some value $z = z_1$ farther from a than α is. By the last theorem, the series must converge at all points nearer to a than z_1 is, and, moreover, the sum must be analytic at every such point But this clearly contradicts the hypothesis that α is a singular point of $f(z)$, and, thus, we have established the following theorem.

THEOREM 4

It is impossible for the Taylor's series of a function $f(z)$ to converge outside the circle whose center is the point of expansion $z = a$ and whose radius is the distance from a to the nearest singular point of $f(z)$

The notion of the circle of convergence is often useful in determining the interval of convergence of a series arising as the expansion of a function of a real variable. To illustrate, consider

$$f(z) = \frac{1}{1 + z^2} = 1 - z^2 + z^4 - z^6 +$$

This will converge throughout the interior of the largest circle around the origin in which $f(z)$ is analytic. Now, by inspection, $f(z)$ is undefined at $z = \pm i$, and even though one may be concerned solely with real values of z [for which $1/(1 + x^2)$ is everywhere infinitely differentiable], these singularities in the complex plane set an inescapable limit to the interval of convergence on the x-axis. We can, in fact, have convergence around $x = a$ on the real axis only over the horizontal diameter of the circle of convergence in the complex plane

As an application of Taylor's expansion, we shall conclude this section by establishing the simple but important result known as the **theorem of Liouville:**[*]

THEOREM 5

If $f(z)$ is bounded and analytic for all values of z, then $f(z)$ is a constant.

PROOF To prove this, we observe first that since $f(z)$ is everywhere analytic, it possesses a power series expansion around the origin

$$f(z) = f(0) + f'(0)z + \cdots + \frac{f^{(n)}(0)}{n!} z^n + \cdots$$

which converges and represents it for all values of z. Now, if C is any circle having the origin as center, it follows from Cauchy's inequality (Theorem 11, Sec. 14.8) that

$$|f^{(n)}(0)| \leq \frac{n! M_C}{r^n}$$

where M_C is the maximum value of $|f(z)|$ on C and r is the radius of C. Hence, for the coefficient of z^n in the expansion of $f(z)$, we have

$$\left|\frac{f^{(n)}(0)}{n!}\right| \leq \frac{M_C}{r^n} \leq \frac{M}{r^n}$$

where M, the bound on $|f(z)|$ for all values of z, which exists by hypothesis, is independent of r. Since r can be taken arbitrarily large, it follows, therefore, that the coefficient of z^n is zero for $n = 1, 2, 3, \ldots$. In other words, for all values of z,

$$f(z) = f(0)$$

which proves the theorem.

A function which is analytic for all values of z is called an **entire function** or an **integral function,** and Liouville's theorem thus states that *any entire function which is bounded for all values of z is necessarily a constant.*

EXERCISES

1 Expand $f(z) = (z - 1)/(z + 1)$ in a Taylor series (a) about the point $z = 0$ and (b) about the point $z = 1$. Determine the region of convergence in each case.

2 Expand $f(z) = \cosh z$ in a Taylor series about the point $z = i\pi$. What is the region of convergence of the resulting series?

3 Expand $f(z) = z/(z + 1)(z + 2)$ in a Taylor's series (a) about the point $z = 0$ and (b) about the point $z = 2$. Determine the region of convergence in each case.

Without obtaining the series, determine the radius of convergence of each of the following expansions:

4 $\tan z$ around $z = 0$

5 $\mathrm{Tan}^{-1} z$ around $z = 1$

6 $1/(e^z - 1)$ around $z = 4i$

7 $x/(x^2 + 2x + 5)$ around $x = 1$

8 Prove that every polynomial equation $P(z) = 0$ has at least one root. [Hint: Assume the contrary and apply Liouville's theorem to the function $f(z) = 1/P(z)$.]

9 Prove that, if the Taylor expansion of a function around a given point exists, it is unique.

---

[*] Named for the French mathematician Joseph Liouville (1809–1882), but actually due to Cauchy.

10 Prove that, if $\sum_{n=0}^{\infty} a_n z^n$ converges absolutely at one point on its circle of convergence, then it converges absolutely and uniformly in the closed region bounded by its circle of convergence

15.3

Laurent's expansion

In many applications it is necessary to expand functions around points at which or in the neighborhood of which the functions are not analytic. The method of Taylor's series is obviously inapplicable in such cases, and a new type of series known as **Laurent's expansion*** is required This furnishes us with a representation which is valid in the annular ring bounded by two concentric circles, provided that the function being expanded is analytic everywhere on and between the two circles As in the case of Taylor's series, the function may have singular points outside the larger circle, and, as the essentially new feature, it may also have singular points within the inner circle The price we pay for this is that negative as well as positive powers of $z - a$ now appear in the expansion and that the coefficients, even of the positive powers of $z - a$, cannot be expressed in terms of the evaluated derivatives of the function The precise result is given by the following theorem

THEOREM 1

If $f(z)$ is analytic throughout the closed region R bounded by two concentric circles, then at any point in the annular ring bounded by the circles, $f(z)$ can be represented by the series

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z - a)^n$$

where a is the common center of the circles and

$$a_n = \frac{1}{2\pi i} \int_C \frac{f(t)\,dt}{(t - a)^{n+1}}$$

each integral being taken in the counterclockwise sense around any curve C lying in the annulus and encircling its inner boundary

PROOF Let z be an arbitrary point of the given annulus. Then according to Cauchy's integral formula we can write

$$\begin{aligned} f(z) &= \frac{1}{2\pi i} \int_{C_2+C_1} \frac{f(t)\,dt}{t - z} \\ &= \frac{1}{2\pi i} \int_{C_2} \frac{f(t)\,dt}{t - z} + \frac{1}{2\pi i} \int_{C_1} \frac{f(t)\,dt}{t - z} \end{aligned}$$

where C_2 is traversed in the counterclockwise direction and C_1 is traversed in

---

* Named for the French mathematician Hermann Laurent (1841–1908)

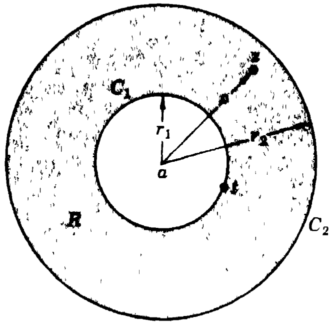


FIGURE 15.3
The circles C_1 and C_2 used in the derivation of Laurent's expansion

the clockwise direction, in order that the entire integration shall be in the positive direction (Fig 15.3) Reversing the sign of the integral around C_1 and also changing the direction of integration from clockwise to counterclockwise, we can write

$$
\begin{aligned}
f(z) &= \frac{1}{2\pi i}\int_{C_2}\frac{f(t)\,dt}{t-z} - \frac{1}{2\pi i}\int_{C_1}\frac{f(t)\,dt}{t-z} \\
&= \frac{1}{2\pi i}\int_{C_2}\frac{f(t)}{t-a}\cdot\frac{1}{1-(z-a)/(t-a)}\,dt \\
&\qquad + \frac{1}{2\pi i}\int_{C_1}\frac{f(t)}{z-a}\,\frac{1}{1-(t-a)/(z-a)}\,dt
\end{aligned}
$$

Now, in each of these integrals let us apply the identity

$$
\frac{1}{1-u} = 1 + u + u^2 + \cdots + \frac{u^n}{1-u}
$$

to the last factor. Then,

$$
\begin{aligned}
f(z) &= \frac{1}{2\pi i}\int_{C_2}\frac{f(t)}{t-a}\left[1 + \frac{z-a}{t-a} + \cdots + \left(\frac{z-a}{t-a}\right)^{n-1}\right. \\
&\qquad\qquad \left. + \frac{(z-a)^n/(t-a)^n}{1-(z-a)/(t-a)}\right]dt \\
&\quad + \frac{1}{2\pi i}\int_{C_1}\frac{f(t)}{z-a}\left[1 + \frac{t-a}{z-a} + \cdots + \left(\frac{t-a}{z-a}\right)^{n-1}\right. \\
&\qquad\qquad \left. + \frac{(t-a)^n/(z-a)^n}{1-(t-a)/(z-a)}\right]dt \\
&= \frac{1}{2\pi i}\int_{C_2}\frac{f(t)\,dt}{t-a} + \frac{z-a}{2\pi i}\int_{C_2}\frac{f(t)\,dt}{(t-a)^2} + \\
&\qquad \cdots + \frac{(z-a)^{n-1}}{2\pi i}\int_{C_2}\frac{f(t)\,dt}{(t-a)^n} + R_{n2} \\
&\quad + \frac{1}{2\pi i(z-a)}\int_{C_1} f(t)\,dt + \frac{1}{2\pi i(z-a)^2}\int_{C_1}(t-a)f(t)\,dt + \\
&\qquad \cdots + \frac{1}{2\pi i(z-a)^n}\int_{C_1}(t-a)^{n-1}f(t)\,dt + R_{n1}
\end{aligned}
$$

where
$$
R_{n2} = \frac{(z-a)^n}{2\pi i}\int_{C_2}\frac{f(t)\,dt}{(t-a)^n(t-z)}
$$

$$
R_{n1} = \frac{1}{2\pi i(z-a)^n}\int_{C_1}\frac{(t-a)^n f(t)\,dt}{z-t}
$$

The truth of the theorem will be established if we can show that

$$\lim_{n\to\infty} R_{n2} = 0 \qquad \text{and} \qquad \lim_{n\to\infty} R_{n1} = 0$$

The proof of the first of these equations we can pass over without comment, because it was given in complete detail in the derivation of Taylor's series in Sec. 15.2. To prove the second, we note that, for values of t on C_1 (Fig. 15.3),

$$\begin{aligned} |t-a| &= r_1 \\ |z-a| &= \rho \qquad \text{say, where } \rho > r_1 \\ |z-t| &= |(z-a)-(t-a)| \geqq \rho - r_1 \end{aligned}$$

and

$$|f(t)| \leqq M$$

where M is the maximum of $|f(z)|$ on C_1 Thus,

$$\begin{aligned} |R_{n1}| &= \left| \frac{1}{2\pi i(z-a)^n} \int_{C_1} \frac{(t-a)^n f(t)\,dt}{z-t} \right| \\ &\leqq \frac{1}{|2\pi i|\,|z-a|^n} \int_{C_1} \frac{|t-a|^n |f(t)|\,|dt|}{|z-t|} \\ &\leqq \frac{r_1{}^n M}{2\pi\rho^n(\rho-r_1)} \int_{C_1} |dt| \\ &= \frac{M}{2\pi}\left(\frac{r_1}{\rho}\right)^n \frac{2\pi r_1}{\rho-r_1} \\ &= M\left(\frac{r_1}{\rho}\right)^n \frac{r_1}{\rho-r_1} \end{aligned}$$

Since $0 < r_1/\rho < 1$, the last expression approaches zero as n becomes infinite Hence, $\lim_{n\to\infty} R_{n1} = 0$; and thus we have

$$\begin{aligned} f(z) = \frac{1}{2\pi i}\int_{C_1}\frac{f(t)\,dt}{t-a} &+ \left[\frac{1}{2\pi i}\int_{C_1}\frac{f(t)\,dt}{(t-a)^2}\right](z-a) \\ &+ \left[\frac{1}{2\pi i}\int_{C_1}\frac{f(t)\,dt}{(t-a)^3}\right](z-a)^2 + \cdots \\ + \left[\frac{1}{2\pi i}\int_{C_1} f(t)\,dt\right]\frac{1}{z-a} &+ \left[\frac{1}{2\pi i}\int_{C_1}(t-a)f(t)\,dt\right]\frac{1}{(z-a)^2} + \cdots \end{aligned}$$

Since $f(z)$ is analytic throughout the region between C_1 and C_2, the paths of integration C_1 and C_2 can be replaced by any other curve C within this region and encircling C_1. The resulting integrals are precisely the coefficients a_n described by the theorem; hence, our proof is complete

It should be noted that the coefficients of the positive powers of $z - a$ in Laurent's expansion, although identical in form with the integrals of Theorem 9, Sec 14 8, *cannot* be replaced by the derivative expressions

$$\frac{f^{(n)}(a)}{n!}$$

as they were in the derivation of Taylor's series, since $f(z)$ is not analytic throughout the entire interior of C_2 (or C), and, hence, Cauchy's generalized integral formula cannot be applied. Specifically, $f(z)$ may have many points of nonanalyticity within C_1 and, therefore, within C_2 (or C).

In many instances the Laurent expansion of a function is found not through the use of the last theorem, but rather by algebraic manipulations suggested by the nature of the function. In particular, in dealing with quotients of polynomials, it is often advantageous to express them in terms of partial fractions and then expand the various denominators in series of the appropriate form through the use of the binomial expansion, which we list here for reference:

THEOREM 2

The expansion

$$(s + t)^n = s^n + ns^{n-1}t + \frac{n(n-1)}{2!}s^{n-2}t^2 + \frac{n(n-1)(n-2)}{3!}s^{n-3}t^3 + \cdots$$

is valid for all values of n if $|s| > |t|$ If $|s| \leq |t|$ the expansion is valid only if n is a nonnegative integer.

That such procedures are correct follows from the fact that *the Laurent expansion of a function over a given annulus is unique.* In other words, if an expansion of the Laurent type is found by any process, it must be *the* Laurent expansion

EXAMPLE 1

Find the Laurent expansion of the function $f(z) = (7z - 2)/(z + 1)z(z - 2)$ in the annulus

$$1 < |z + 1| < 3$$

As a preliminary step it is convenient to apply the method of partial fractions to $f(z)$ and express it in the form

$$f(z) = \frac{-3}{z+1} + \frac{1}{z} + \frac{2}{z-2}$$

Now, after suitable rearrangement, these terms can be expanded into infinite series by means of Theorem 2 and added to give the required expansion for $f(z)$

To do this, we observe that since the center of the given annulus is $z = -1$, the series we are seeking must be one involving powers of $z + 1$ Hence, we modify the second and third terms in the partial-fraction representation of $f(z)$ so that z will appear in the combination $z + 1$ This gives us the equivalent expression

$$\begin{aligned} f(z) &= \frac{-3}{z+1} + \frac{1}{(z+1) - 1} + \frac{2}{(z+1) - 3} \\ &= -3(z+1)^{-1} + [(z+1) - 1]^{-1} + 2[(z+1) - 3]^{-1} \end{aligned}$$

However, according to Theorem 2, the series for $[(z + 1) - 3]^{-1}$ will converge only where $|z + 1| > 3$, whereas we require an expansion valid for $|z + 1| < 3$ Hence, we rewrite this term in the other order, $[-3 + (z + 1)]^{-1}$, before expanding it. Now we can apply Theorem 2, obtaining

$$\begin{aligned} f(z) &= -3(z+1)^{-1} + [(z+1) - 1]^{-1} + 2[-3 + (z+1)]^{-1} \\ &= -3(z+1)^{\ 1} + [(z+1)^{-1} + (z+1)^{-2} + (\ +1)^{-3} + \quad] \\ &\qquad + 2\left[-\frac{1}{3} - \frac{z+1}{9} - \frac{(z+1)^2}{27} - \frac{(z+1)^3}{81} - \quad \right] \\ &= \quad + (z+1)^{-2} - 2(z+1)^{-1} - \tfrac{2}{3} - \tfrac{2}{9}(z+1) \\ &\qquad - \tfrac{2}{27}(z+1)^2 - \cdots \qquad 1 < |z+1| < 3 \end{aligned}$$

It is important to note that $f(z)$ has two other Laurent expansions around the point $z = -1$. One is valid in the annular region between a circle of arbitrarily small radius around $z = -1$ and a circle of unit radius around $z = -1$. The other is valid in the region exterior to a circle of radius 3 around $z = -1$ (Fig. 15.4). Each of these can be found, as above, by suitably rearranging the terms in the partial-fraction representation of $f(z)$ and then expanding these terms by means of Theorem 2. Thus, in the innermost region we have

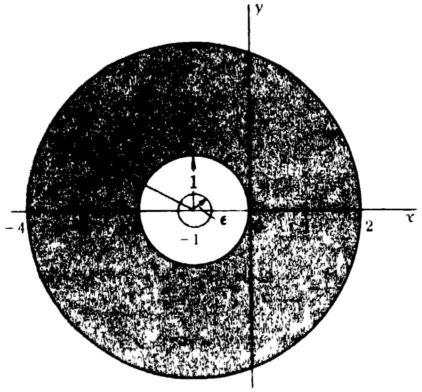


FIGURE 15.4 The regions of validity of the three Laurent expansions of $(7z - 2)/[(z + 1)z(z - 2)]$ around $z = -1$

$$\begin{aligned} f(z) &= -3(z+1)^{-1} + [-1 + (z+1)]^{-1} + 2[-3 + (z+1)]^{-1} \\ &= -3(z+1)^{-1} + [-1 - (z+1) - (z+1)^2 - (z+1)^3 - \quad] \\ &\qquad + 2\left[-\frac{1}{3} - \frac{z+1}{9} - \frac{(z+1)^2}{27} - \frac{(z+1)^3}{81} - \quad \right] \\ &= -3(z+1)^{-1} - 5/3 - {}^{11}\!/_{9}(z+1) - {}^{29}\!/_{27}(z+1)^2 \\ &\qquad - {}^{83}\!/_{81}(z+1)^3 - \qquad 0 < |z+1| < 1 \end{aligned}$$

Similarly, in the outermost region we have

$$\begin{aligned} f(z) &= -3(z+1)^{-1} + [(z+1) - 1]^{-1} + 2[(z+1) - 3]^{-1} \\ &= -3(z+1)^{-1} + [(z+1)^{-1} + (z+1)^{-2} + (z+1)^{-3} + \quad] \\ &\qquad + 2[(z+1)^{-1} + 3(z+1)^{-2} + 9(z+1)^{-3} + \quad] \\ &= \qquad + 19(z+1)^{-3} + 7(z+1)^{-2} \qquad |z+1| > 3 \end{aligned}$$

Incidentally, the fact that we have obtained these Laurent expansions without using the general theory means that we can evaluate the integrals in the coefficient formulas by comparing them with the numerical values of the coefficients we have found by independent means. For instance, in the first expansion the coefficient of $(z + 1)^{-1}$ is -2. On the other hand, according to the theory of Laurent's expansion, the coefficient of this term is

$$a_{-1} = \frac{1}{2\pi i}\int_C f(z)\,dz = \frac{1}{2\pi i}\int_C \frac{7z - 2}{(z+1)z(z-2)}\,dz$$

where C is any closed curve lying in the interior of the circle $|z + 1| = 3$ and enclosing the circle $|z + 1| = 1$. Thus, although we have done nothing resembling an integration, we have nonetheless shown that

$$\frac{1}{2\pi i}\int_C \frac{7z - 2}{(z+1)z(z-2)}\,dz = -2 \qquad \text{or} \qquad \int_C \frac{7z - 2}{(z+1)z(z-2)}\,dz = -4\pi i$$

a result, incidentally, which could not have been obtained by a direct application of Cauchy's integral formula, as in Example 2, Sec. 14.8.

EXERCISES

1 Expand $f(z) = 1/(z-1)(z-2)$

a For $|z| < 1$ **b** For $1 < |z| < 2$ **c** For $2 < |z|$
d For $0 < |z-1| < 1$ **e** For $|z-1| > 1$
f For $0 < |z-2| < 1$ **g** For $|z-2| > 1$

2 Obtain two distinct Laurent expansions for $f(z) = (3z+1)/(z^2-1)$ around $z = 1$, and tell where each converges

3 Expand $f(z) = 1/z^2(z-i)$ in two different Laurent expansions around $z = i$, and tell where each converges

4 Construct all the Laurent expansions of $f(z) = 1/z(z-1)(z-2)$ around $z = -1$, and tell where each converges

5 Find the value of $\int_C f(z)\,dz$ if C is the circle $|z| = 3$ and $f(z)$ is

a $\dfrac{1}{z(z+2)}$ **b** $\dfrac{z+2}{z(z+1)}$ **c** $\dfrac{1}{(z+1)^2}$

d $\dfrac{1}{z(z+1)^2}$ **e** $\dfrac{z}{(z+1)(z+2)}$ **f** $\dfrac{1}{z(z+1)(z+4)}$

6 If k is a real number such that $k^2 < 1$, prove that

$$\sum_{n=0}^{\infty} k^n \sin(n+1)\theta = \frac{\sin\theta}{1 - 2k\cos\theta + k^2}$$

$$\sum_{n=0}^{\infty} k^n \cos(n+1)\theta = \frac{\cos\theta - k}{1 - 2k\cos\theta + k^2}$$

[Hint Expand $(z-k)^{-1}$ for $|z| > k$, set $z = e^{i\theta}$, and equate real and imaginary components in the resulting expression]

7 Criticize the following argument Since (by long division, for instance)

$$\frac{z}{1-z} = z + z^2 + z^3 + z^4 + \cdots \qquad \text{and} \qquad \frac{z}{z-1} = 1 + \frac{1}{z} + \frac{1}{z^2} + \frac{1}{z^3} + \cdots$$

and since $\dfrac{z}{1-z} + \dfrac{z}{z-1} = 0$

therefore, by adding these two series we obtain

$$\cdots + \frac{1}{z^3} + \frac{1}{z^2} + \frac{1}{z} + 1 + z + z^2 + z^3 + z^4 + \cdots = 0$$

8 Criticize the following argument The series

$$\frac{1}{z} + 1 + z + z^2 + z^3 + z^4 + \cdots$$

converges to the sum $S(z) = \dfrac{1}{z(1-z)}$ for all values of z such that $|z| < 1$, *including* $z = 0$, since

$$\begin{aligned}|S(z) - S_n(z)| &= \left|\frac{1}{z(1-z)} - \left(\frac{1}{z} + 1 + z + \cdots + z^{n-2}\right)\right| \\ &= \left|\frac{1}{z} + \frac{1}{1-z} - \frac{1}{z} - 1 - z - \cdots - z^{n-2}\right| \\ &= \left|\frac{1}{1-z} - 1 - z - \cdots - z^{n-2}\right| \\ &= \left|\frac{z^{n-1}}{1-z}\right|\end{aligned}$$

and this expression clearly approaches 0 as n becomes infinite for *all* values of z such that $|z| < 1$.

9 **a** Show that the Laurent expansion of $f(z) = \sinh\left(z + \frac{1}{z}\right)$ in powers of z is $f(z) = \sum_{n=-\infty}^{\infty} a_n z^n$, where

$$a_n = \frac{1}{2\pi}\int_0^{2\pi} \cos n\theta \sinh (2 \cos \theta)\, d\theta$$

(Hint In the formula for a_n provided by Theorem 1, take the curve C to be the circle $|z| = 1$ On this circle let the variable of integration t be taken in the form $t = e^{i\theta}$ Finally, verify that the imaginary part of the integral for a_n is equal to zero)

b Show that the coefficients in the Laurent expansion of $f(z) = \sin\left(z + \frac{1}{z}\right)$ in powers of z are given by the formula $a_n = \frac{1}{2\pi}\int_0^{2\pi} \cos n\theta \sin (2 \cos \theta)\, d\theta$

c What are the coefficients in the Laurent expansion of $f(z) = \cosh\left(z + \frac{1}{z}\right)$ in powers of z?

d What are the coefficients in the Laurent expansion of $f(z) = \cos\left(z + \frac{1}{z}\right)$ in powers of z?

10 Let $f(z) = f(re^{i\theta}) = F(r,\theta)$ be a function which is analytic in some annulus having the origin as center and containing the circle $|z| = 1$ Taking this circle as the curve C in the formula for a_n in the Laurent expansion of $f(z)$, show that

$$F(r,\theta) = \frac{1}{2\pi}\int_0^{2\pi} F(1,\phi)\, d\phi + \frac{1}{\pi}\sum^{\infty} \int_0^{2\pi} F(1,\phi) \cos n(\theta - \phi)\, d\phi$$

CHAPTER SIXTEEN

The Theory of Residues

16.1

The residue theorem

In Sec. 14.6 we defined a singular point of a function $f(z)$ as a point where $f(z)$ is not analytic but in every neighborhood of which there are points where $f(z)$ is analytic. If $z = a$ is a singular point of the function $f(z)$, but if there exists a neighborhood of a in which there are no other singular points of $f(z)$, then $z = a$ is called an **isolated singular point.** Clearly, if $z = a$ is an isolated singularity of $f(z)$, then $f(z)$ will possess a Laurent expansion around $z = a$ which will be valid in the interior of an annulus whose outer radius is the distance from a to the nearest of the other singular points of $f(z)$ and whose inner radius can be taken arbitrarily small.

If the Laurent expansion of $f(z)$ in the neighborhood of an isolated singular point $z = a$ contains only a finite number of negative powers of $z - a$, then $z = a$ is called a **pole** of $f(z)$. If $(z - a)^{-m}$ is the highest negative power in the expansion, the pole is said to be of **order** m, and the sum of all the terms containing negative powers, namely,

$$\frac{a_{-m}}{(z-a)^m} + \cdots + \frac{a_{-2}}{(z-a)^2} + \frac{a_{-1}}{z-a}$$

is called the **principal part** of $f(z)$ at $z = a$. If the Laurent expansion of $f(z)$ in the neighborhood of an isolated singular point $z = a$ contains infinitely many negative powers of $z - a$, then $z = a$ is called an **essential singularity** of $f(z)$. For instance, since

$$\begin{aligned}\frac{1}{z(z-1)^2} &= \frac{[1 + (z-1)]^{-1}}{(z-1)^2} \\ &= \frac{1}{(z-1)^2} - \frac{1}{z-1} + 1 - (z-1) + \cdots \\ &\qquad\qquad 0 < |z-1| < 1\end{aligned}$$

this function has a pole of order 2 at $z = 1$, and its principal part there is

$$\frac{1}{(z-1)^2} - \frac{1}{z-1}\dagger$$

On the other hand, since $e^{1/z}$ is represented for all values of z except $z = 0$ by the series

$$e^{1/z} = 1 + \frac{1}{z} + \frac{1}{2!z^2} + \frac{1}{3!z^3} + \frac{1}{4!z^4} + \cdots$$

it has an essential singularity at the origin

In passing, we note that, if the terms in the expansion of $f(z)$ around a pole of order m, say $z = a$, are put over a common denominator, $f(z)$ will contain the factor $1/(z-a)^m$ Conversely, if a function $f(z)$ is expressed as a fraction in lowest terms, then the presence of a factor of the form $(z-a)^m$ in the denominator implies that $f(z)$ has a pole of the mth order at $z = a$ In most applications this is the way in which the poles of a function are found

As we suggested at the end of the last chapter, the coefficient a_{-1} of the term $(z-a)^{-1}$ in the Laurent expansion of a function $f(z)$ is of great importance because of its connection with the integral of the function, through the formula

$$a_{-1} = \frac{1}{2\pi i}\int_C f(z)\,dz$$

In particular, the coefficient of $(z-a)^{-1}$ in the expansion of $f(z)$ *in the neighborhood of an isolated singular point* is called the **residue** of $f(z)$ at that point

Now consider a simple closed curve C containing in its interior a number of isolated singularities of a function $f(z)$ If around each singular point we draw a circle so small that it encloses no other singular points (Fig 16 1), these circles, together with the curve C, form the boundary of a multiply connected region in which $f(z)$ is everywhere analytic and to which Cauchy's theorem can, therefore, be applied This gives

$$\frac{1}{2\pi i}\int_C f(z)\,dz + \frac{1}{2\pi i}\int_{C_1} f(z)\,dz + \cdots + \frac{1}{2\pi i}\int_{C_n} f(z)\,dz = 0$$

If we reverse the direction of integration around each of the circles and change the sign of each integral to compensate, this

---

† It should be noted that, although we can also write

$$\frac{1}{z(z-1)^2} = \frac{[(z-1)+1]^{-1}}{(z-1)^2}$$
$$= \cdots + \frac{1}{(z-1)^5} - \frac{1}{(z-1)^4} + \frac{1}{(z-1)^3} \qquad |z-1| > 1$$

the fact that this expansion contains infinitely many negative powers of $z - 1$ does not contradict our observation that $1/z(z-1)^2$ has a pole of order 2 at $z = 1$ For this series is valid only *outside* the circle $|z-1| = 1$, whereas the presence of poles and essential singularities is determined by the structure of the particular Laurent expansion which is valid in the *innermost* annulus, or deleted neighborhood, of the point in question

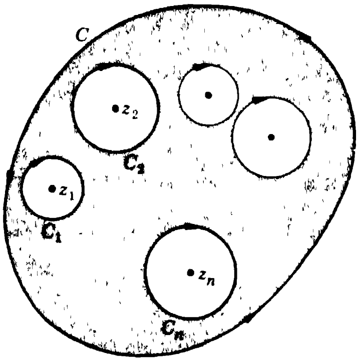


FIGURE 16.1 The circles C_1, C_2, , C_n enclosing, respectively, the singular points z_1, z_2, , z_n within a simple closed curve

can be written

$$\frac{1}{2\pi i}\int_C f(z)\,dz = \frac{1}{2\pi i}\int_{C_1} f(z)\,dz + \qquad + \frac{1}{2\pi i}\int_{C_n} f(z)\,dz$$

where all the integrals are now to be taken in the counterclockwise sense But the integrals on the right are, by definition, just the residues of $f(z)$ at the various isolated singularities within C. Hence we have established the important **residue theorem**:

THEOREM 1

If C is a closed curve and if $f(z)$ is analytic within and on C except at a finite number of singular points in the interior of C, then

$$\int_C f(z)\,dz = 2\pi i(r_1 + r_2 + \quad\cdot\quad + r_n)$$

where r_1, r_2, , r_n are the residues of $f(z)$ at its singular points within C

EXAMPLE 1

What is the integral of

$$f(z) = \frac{-3z + 4}{z(z-1)(z-2)}$$

around the circle $|z| = \tfrac{3}{2}$?

In this case, although there are three singular points of the function, namely, the three first-order poles at $z = 0$, $z = 1$, and $z = 2$, only $z = 0$ and $z = 1$ lie within the path of integration Hence, the core of the problem is to find the residues of $f(z)$ at these two points

To do this, it is natural to begin by constructing the partial-fraction representation of $f(z)$, namely,

$$f(z) = \frac{2}{z} - \frac{1}{z-1} - \frac{1}{z-2}$$

Then, in the neighborhood of $z = 0$, we can write

$$\begin{aligned} f(z) &= \frac{2}{z} + (1-z)^{-1} + (2-z)^{-1} \\ &= \frac{2}{z} + (1 + z + z^2 + \qquad) + \left(\frac{1}{2} + \frac{z}{4} + \frac{z^2}{8} + \cdot\cdot\right) \\ &= \frac{2}{z} + \frac{3}{2} + \frac{5}{4}z + \frac{9}{8}z^2 + \quad\cdot \end{aligned}$$

Hence, the residue of $f(z)$ at $z = 0$, i.e., the coefficient of the term $1/z$ in the last expansion, is 2.[†] Also, in the neighborhood of $z = 1$, we have

$$\begin{aligned} f(z) &= 2[1 + (z-1)]^{-1} - \frac{1}{z-1} + [1 - (z-1)]^{-1} \\ &= 2[1 - (z-1) + (z-1)^2 - \cdots] - \frac{1}{z-1} + [1 + (z-1) + (z-1)^2 + \cdots] \\ &= \frac{-1}{z-1} + 3 - (z-1) + 3(z-1)^2 - \cdots \end{aligned}$$

Hence, the residue of $f(z)$ at $z = 1$ is -1. Therefore, according to the residue theorem,

$$\int_C \frac{-3z+4}{z(z-1)(z-2)}\,dz = 2\pi i[(2) + (-1)] = 2\pi i$$

The determination of residues by the use of series expansions, in the manner we have just illustrated, is often tedious and sometimes very difficult. Hence, it is desirable to have a simpler alternative procedure. Such a process is provided by the following considerations. Suppose first that $f(z)$ has a simple, or first-order, pole at $z = a$. It follows, then, that we can write

$$f(z) = \frac{a_{-1}}{z-a} + a_0 + a_1(z-a) + \cdots$$

If we multiply this identity by $z - a$, we get

$$(z-a)f(z) = a_{-1} + a_0(z-a) + a_1(z-a)^2 + \cdots$$

Now, if we let z approach a, we obtain for the residue

(1) $$a_{-1} = \lim_{z\to a}\,[(z-a)f(z)]$$

If $f(z)$ has a second-order pole at $z = a$, then

$$f(z) = \frac{a_{-2}}{(z-a)^2} + \frac{a_{-1}}{z-a} + a_0 + a_1(z-a) + a_2(z-a)^2 + \cdots$$

Now, to obtain the residue a_{-1}, we must multiply this identity by $(z-a)^2$ and then differentiate with respect to z before we let z approach a. The result this time is

(2) $$a_{-1} = \lim_{z\to a}\frac{d}{dz}\,[(z-a)^2 f(z)]$$

The same procedure can be extended to poles of higher order leading to the formula contained in the following theorem:

THEOREM 2

If $f(z)$ has a pole of order m at $z = a$, then the residue of $f(z)$ at $z = a$ is

$$a_{-1} = \frac{1}{(m-1)!}\lim_{z\to a}\frac{d^{m-1}}{dz^{m-1}}\,[(z-a)^m f(z)]$$

[†] Since $1/(z-1)$ and $1/(z-2)$ are both analytic in the neighborhood of $z = 0$, it is evident in advance that their expansions around $z = 0$ will be, not Laurent, but Taylor series and, hence, will contain no negative powers of z. Thus neither of these terms can contribute to the residue of $f(z)$ at $z = 0$, and so it is actually unnecessary to obtain their expansions. The same thing is true of the terms $2/z$ and $1/(z-2)$ around $z = 1$.

In many problems the order of the pole at $z = a$ will not be known in advance In such cases it is still possible to apply Theorem 2 by taking $m = 1, 2, 3, \ldots$, in turn, until for the *first* time a finite limit is obtained for a_{-1} The value of m for which this occurs is the order of the pole, and the value of a_{-1} thus determined is the residue If $f(z)$ has an essential singularity at $z = a$, however, this process fails, and the residue cannot be determined by means of Theorem 2.

EXAMPLE 2

What is the residue of $f(z) = (1 + z)/(1 - \cos z)$ at the origin?

Here the order of the pole is unknown, so it appears that we may have to proceed tentatively, trying $m = 1, 2, 3, \ldots$, in turn, until for the first time we obtain a finite value for the residue a_{-1} However, if we replace $\cos z$ by its Maclaurin expansion, we obtain for $f(z)$ the expression

$$\frac{1 + z}{1 - \left(1 - \dfrac{z^2}{2} + \dfrac{z^4}{24} - \cdots\right)} = \frac{2(1 + z)}{z^2\left(1 - \dfrac{z^2}{12} + \cdots\right)}$$

and the factor z^2 in the denominator now identifies the pole as of the second order. Hence, applying Theorem 2 with $m = 2$, we have

$$\begin{aligned} a_{-1} &= \lim_{z\to 0} \frac{d}{dz}\left[z^2 \frac{2(1 + z)}{z^2\left(1 - \dfrac{z^2}{12} + \cdots\right)}\right] \\ &= \lim_{z\to 0} 2\,\frac{\left(1 - \dfrac{z^2}{12} + \cdots\right) - (1 + z)\left(-\dfrac{z}{6} + \cdots\right)}{\left(1 - \dfrac{z^2}{12} + \cdots\right)^2} \\ &= 2 \end{aligned}$$

EXERCISES

1 Find the residue of $f(z) = z/(z^2 + 1)$ (a) at $z = i$ and (b) at $z = -i$

2 Find the residue of $f(z) = (z + 1)/z^2(z - 2)$ (a) at $z = 0$ and (b) at $z = 2$

3 Find the residue of $f(z) = z/(z^2 + 2z + 5)$ at each of its poles

4 What is the residue of $f(z) = 1/(z + 1)^3$ at $z = -1$?

5 What is the residue of $f(z) = \tan z$ at $z = \pi/2$?

6 What is the residue of $f(z) = z/(\cosh z - \cos z)$ at $z = 0$?

7 What is the residue of $f(z) = 1/(z - \sin z)$ at $z = 0$?

8 What is the residue of $f(z) = 1/(e^z - 1)$ at $z = 0$?

9 If C is the circle $|z| = 4$, evaluate $\int_C f(z)\, dz$ for the functions

a $f(z) = \dfrac{z}{z^2 - 1}$ **b** $f(z) = \dfrac{z + 1}{z^2(z + 2)}$

c $f(z) = \dfrac{1}{z(z - 2)^3}$ **d** $f(z) = \dfrac{z^2}{(z^2 + 3z + 2)^2}$

e $f(z) = \dfrac{1}{z^2 + z + 1}$ **f** $f(z) = \dfrac{1}{z(z^2 + 6z + 4)}$

10 If C is the circle $|z| = 2$, evaluate $\int_C f(z)\,dz$ for the functions

a $f(z) = \tan z$ b $f(z) = \dfrac{1}{z \sin z}$

c $f(z) = \dfrac{1}{z^2 \sin z}$ d $f(z) = \dfrac{e^{-z}}{z^2}$

e $f(z) = ze^{1/z}$ f $f(z) = \dfrac{z}{\cos z}$

16.2

The evaluation of real definite integrals

There are several large and important classes of real definite integrals whose evaluation by the theory of residues can be made a routine matter. The results in question are contained in the next three theorems.

THEOREM 1

If $R(\cos\theta, \sin\theta)$ is a rational function of $\cos\theta$ and $\sin\theta$ which is finite on the closed interval $0 \leqq \theta \leqq 2\pi$ and if $f(z)$ is the function obtained from R by the substitutions

$$\cos\theta = \frac{z + z^{-1}}{2} \qquad \sin\theta = \frac{z - z^{-1}}{2i}$$

then $\int_0^{2\pi} R(\cos\theta, \sin\theta)\,d\theta$ is equal to $2\pi i$ times the sum of the residues of the function $\dfrac{f(z)}{iz}$ at such of its poles as lie within the unit circle $|z| = 1$.

PROOF As a first step, let us transform the given integral by means of the substitution $z = e^{i\theta}$, according to which

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} = \frac{z + z^{-1}}{2} \qquad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} = \frac{z - z^{-1}}{2i} \qquad d\theta = \frac{dz}{iz}$$

Under this transformation the original integrand becomes a rational function of z, which we call $f(z)$. Furthermore, as θ ranges from 0 to 2π, the relation $z = e^{i\theta}$ shows that z ranges around the unit circle $|z| = 1$. Hence, the transformed integral is

$$\int_C f(z)\,\frac{dz}{iz}$$

where C is the unit circle. By the residue theorem, the value of this integral is $2\pi i$ times the sum of the residues at those poles of its integrand, namely, $f(z)/iz$, which lie within the unit circle. Since this integral is equal to the original one, the theorem is established.

EXAMPLE 1

Evaluate $\displaystyle\int_0^{2\pi} \frac{\cos 2\theta\, d\theta}{1 - 2p\cos\theta + p^2}$ $(-1 < p < 1)$

Since the denominator of the integrand can be written

$$\begin{aligned} 1 - 2p\cos\theta + p^2 &= 1 - 2p + p^2 + 2p - 2p\cos\theta = (1-p)^2 + 2p(1-\cos\theta) \\ &= 1 + 2p + p^2 - 2p - 2p\cos\theta = (1+p)^2 - 2p(1+\cos\theta) \end{aligned}$$

it is clear that it can never vanish for $0 \leqq \theta \leqq 2\pi$ if $-1 < p < 1$ Hence, the preceding theorem is applicable Now

$$\cos 2\theta = \frac{e^{2i\theta} + e^{-2i\theta}}{2} = \frac{z^2 + z^{-2}}{2}$$

and thus the given integral becomes

$$\begin{aligned} \int \frac{z^2 + z^{-2}}{2}\,\frac{1}{1 - 2p(z + z^{-1})/2 + p^2}\,\frac{dz}{iz} &= \int \frac{z^4 + 1}{2z^2}\,\frac{z}{z - pz^2 - p + p^2 z}\,\frac{dz}{iz} \\ &= \int \frac{(1 + z^4)\,dz}{2iz^2(1 - pz)(z - p)} \end{aligned}$$

Of the three poles of the integrand, only the first-order pole at $z = p$ and the second-order pole at $z = 0$ lie within the unit circle For the residue at the former we have

$$\lim_{z \to p}\,(z - p)\,\frac{1 + z^4}{2iz^2(1 - pz)(z - p)} = \frac{1 + p^4}{2ip^2(1 - p^2)}$$

For the residue at the second-order pole $z = 0$, we have

$$\begin{aligned} \lim_{z \to 0} \frac{d}{dz}\left[z^2\,\frac{1 + z^4}{2iz^2(z - pz^2 - p + p^2 z)}\right] &= \lim_{z \to 0} \frac{(z - pz^2 - p + p^2 z)(4z^3) - (1 + z^4)(1 - 2pz + p^2)}{2i(z - pz^2 - p + p^2 z)^2} \\ &= -\frac{1 + p^2}{2ip^2} \end{aligned}$$

By Theorem 1, the value of the integral is therefore

$$2\pi i\left[\frac{1 + p^4}{2ip^2(1 - p^2)} - \frac{1 + p^2}{2ip^2}\right] = \frac{2\pi p^2}{1 - p^2}$$

THEOREM 2

If $Q(z)$ is a function which is analytic in the upper half of the z-plane except at a finite number of poles none of which lies on the real axis and if $zQ(z)$ converges uniformly to zero when $z \to \infty$ through values for which $0 \leqq \arg z \leqq \pi$, then $\int_{-\infty}^{\infty} Q(x)\,dx$ is equal to $2\pi i$ times the sum of the residues at the poles of $Q(z)$ which lie in the upper half plane

PROOF We consider a semicircular contour with center at $z = 0$ and with radius R large enough to include all the poles of $Q(z)$ which lie in the upper half plane (Fig 16 2) Then, by the residue theorem,

$$\int_{C_1 + C_2} Q(z)\,dz = 2\pi i \sum \text{residues of } Q(z) \text{ at all poles within } C_1 + C_2$$

or

$$\int_{-R}^{R} Q(x)\,dx + \int_{C_2} Q(z)\,dz = 2\pi i \sum \text{residues}$$

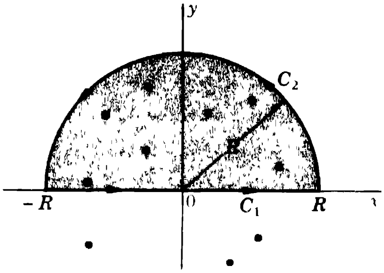


FIGURE 16.2 A semicircular contour enclosing all the poles of a function which lie in the upper half plane.

Hence,

$$(1) \qquad \left| \int_{-R}^{R} Q(x)\,dx - 2\pi i \sum \text{residues} \right| = \left| - \int_{C_2} Q(z)\,dz \right|$$

In the integral on the right, let $z = Re^{i\theta}$, so that $dz = Rie^{i\theta}\,d\theta = iz\,d\theta$. Then

$$\left| - \int_{C_2} Q(z)\,dz \right| = \left| - \int_0^{\pi} Q(z) iz\,d\theta \right| \leqq \int_0^{\pi} |zQ(z)|\,|d\theta|$$

But from the hypothesis that $|zQ(z)|$ converges *uniformly* to zero when $z \to \infty$ and $0 \leqq \arg z \leqq \pi$, it follows that, for any arbitrarily small positive quantity, say ϵ/π, there exists a radius R_0 such that

$$|zQ(z)| < \frac{\epsilon}{\pi}$$

for all values of z on C_2 whenever $R > R_0$. Thus, for $R > R_0$,

$$\int_0^{\pi} |zQ(z)|\,|d\theta| < \frac{\epsilon}{\pi} \int_0^{\pi} |d\theta| = \epsilon$$

This, coupled with (1), proves that

$$\lim_{R\to\infty} \int_{-R}^{R} Q(x)\,dx = 2\pi i \sum \text{residues}$$

Since the limit on the left is what we mean by $\int_{-\infty}^{\infty} Q(x)\,dx$,† the theorem is established.

In particular, the quotient of two polynomials $p(x)/q(x)$ automatically satisfies all the hypotheses of the last theorem whenever the degree of the denominator exceeds the degree of the numerator by at least 2. Hence, we have the following highly important corollary:

---

† Actually $\lim_{R\to\infty} \int_{-R}^{R} Q(x)\,dx$ is only the **principal value** of the integral $\int_{-\infty}^{\infty} Q(x)\,dx$, whose correct definition is

$$\lim_{R\to\infty} \int_{-R}^{0} Q(x)\,dx + \lim_{S\to\infty} \int_0^{S} Q(x)\,dx$$

where R and S become infinite independently of each other As the simple function $Q(x) = x$ shows, the principal value of an integral may exist when the integral itself is undefined. However, under the relatively stringent conditions of Theorem 2 the existence of the principal value implies the existence of the integral itself

COROLLARY 1

If $p(x)$ and $q(x)$ are real polynomials such that the degree of $q(x)$ is at least 2 more than the degree of $p(x)$ and if $q(x) = 0$ has no real roots, then

$$\int_{-\infty}^{\infty} \frac{p(x)}{q(x)}\,dx = 2\pi i \sum \text{residues of } \frac{p(z)}{q(z)} \text{ at its poles in the upper half plane}$$

EXAMPLE 2

Evaluate $\displaystyle\int_{-\infty}^{\infty} \frac{x^2\,dx}{(x^2+a^2)(x^2+b^2)}$

This is an integral to which the corollary of Theorem 2 can surely be applied. The only poles of

$$\frac{z^2}{(z^2+a^2)(z^2+b^2)}$$

are at $z = \pm ai$, $\pm bi$. Of these, only $z = ai$ and $z = bi$ lie in the upper half plane. At $z = ai$ the residue is

$$\lim_{z\to ai}\,(z-ai)\frac{z^2}{(z-ai)(z+ai)(z^2+b^2)} = \frac{-a^2}{2ai(b^2-a^2)} = \frac{a}{2i(a^2-b^2)}$$

From symmetry, the residue at $z = bi$ is obviously $b/2i(b^2-a^2)$. Hence, the value of the integral is

$$2\pi i\left[\frac{a}{2i(a^2-b^2)} + \frac{b}{2i(b^2-a^2)}\right] = \frac{\pi}{a+b}$$

If $Q(z)$ satisfies all the hypotheses of Theorem 2, then so does $e^{imz}Q(z)$, provided $m > 0$. For e^{imz} is analytic everywhere, and, under the assumption that $m > 0$, its absolute value is

$$|e^{imz}| = |e^{im(x+iy)}| = |e^{imx}e^{-my}| = e^{-my}$$

which is less than or equal to 1 for all values of y in the upper half plane. Therefore,

$$|e^{imz}zQ(z)| \leqq |zQ(z)|$$

and thus, if the latter converges uniformly to zero when $z \to \infty$ and $0 \leqq \arg z \leqq \pi$, so will the former. Hence, the conclusions of Theorem 2 can be applied equally well to $e^{imz}Q(z)$, and we can write

$$\int_{-\infty}^{\infty} e^{imx}Q(x)\,dx = 2\pi i \sum \text{residues of } e^{imz}Q(z) \text{ at its poles in the upper half plane} \tag{2}$$

Separating the integral in (2) into its real and its imaginary parts and equating these to the corresponding parts of the right-hand side, we obtain the following useful result:

COROLLARY 2

If $Q(z)$ is analytic in the upper half of the z-plane except at a finite number of poles none of which lies on the real axis and if $|zQ(z)|$ converges uniformly to zero when z becomes infinite through the upper half plane, then

$$\int_{-\infty}^{\infty} \cos mx\, Q(x)\, dx = -2\pi \sum \text{imaginary parts of the residues of } e^{imz}Q(z) \text{ at its poles in the upper half plane}$$

$$\int_{-\infty}^{\infty} \sin mx\, Q(x)\, dx = 2\pi \sum \text{real parts of the residues of } e^{imz}Q(z) \text{ at its poles in the upper half plane}$$

EXAMPLE 3

Evaluate $\displaystyle\int_{-\infty}^{\infty} \frac{\cos mx}{1 + x^2}\, dx$.

To do this, we consider the related function $e^{imz}/(1 + z^2)$ The only pole of this function in the upper half plane is $z = i$, and the residue there is

$$\lim_{z \to i} (z - i)\frac{e^{imz}}{(z - i)(z + i)} = \frac{e^{-m}}{2i} = -\frac{ie^{-m}}{2}$$

Hence, by Corollary 2,

$$\int_{-\infty}^{\infty} \frac{\cos mx}{1 + x^2}\, dx = -2\pi\mathscr{I}\left(-\frac{ie^{-m}}{2}\right) = \pi e^{-m}$$

Incidentally, the fact that the residue at $z = i$ is a pure imaginary quantity confirms the observation, obvious from symmetry, that

$$\int_{-\infty}^{\infty} \frac{\sin mx}{1 + x^2}\, dx = 0$$

As a final result on the evaluation of real definite integrals by the method of residues, we have the following theorem, whose proof we omit because of its relative intricacy.*

THEOREM 3

If $Q(z)$ is analytic everywhere in the z-plane except at a finite number of poles none of which lies on the positive half of the real axis and if $|z^aQ(z)|$ converges uniformly to zero when $z \to 0$ and when $z \to \infty$, then

$$\int_0^{\infty} x^{a-1}Q(x)\, dx = \frac{\pi}{\sin a\pi} \sum \text{residues of } (-z)^{a-1}Q(z) \text{ at all its poles}$$

provided that arg z is taken in the interval $(-\pi,\pi)$

In applying this theorem it must be borne in mind that unless a is an integer, $(-z)^{a-1}$ is a multiple-valued function which, according to Eq. (23), Sec 14 7, is to be interpreted as

$$(-z)^{a-1} = e^{(a-1)\ln(-z)} = e^{(a-1)[\ln|z| + i\arg(-z)]} \qquad -\pi < \arg z \leqq \pi$$

EXAMPLE 4

Evaluate $\displaystyle\int_0^{\infty} \frac{x^{a-1}}{1 + x^2}\, dx$, $0 < a < 2$.

For a within the specified range, the conditions of Theorem 3 are fulfilled, hence the given integral is equal to $\pi/(\sin a\pi)$ times the sum of the residues of $(-z)^{a-1}/(1 + z^2)$ at $z = \pm i$ At

---

* See, for instance, E T Whittaker and G N. Watson, "Modern Analysis," p. 117, The Macmillan Company, New York, 1943.

$z = i$ we have for the residue

$$\lim_{z\to i}(z-i)\frac{(-z)^{a-1}}{(z-i)(z+i)} = \frac{(-i)^{a-1}}{2i} = \frac{(e^{-i\pi/2})^{a-1}}{2i} = \frac{e^{-i\pi(a-1)/2}}{2i}$$

At $z = -i$, for the residue we have

$$\lim_{z\to -i}(z+i)\frac{(-z)^{a-1}}{(z+i)(z-i)} = \frac{i^{a-1}}{-2i} = \frac{(e^{i\pi/2})^{a-1}}{-2i} = \frac{e^{i\pi(a-1)/2}}{-2i}$$

The value of the integral is, therefore,

$$\frac{\pi}{\sin a\pi}\cdot\frac{e^{i\pi(a-1)/2} - e^{-i\pi(a-1)/2}}{-2i} = -\frac{\pi}{\sin a\pi}\sin\frac{(a-1)\pi}{2}$$

$$= \frac{\pi}{\sin a\pi}\cos\frac{a\pi}{2} = \frac{\pi}{2\sin(a\pi/2)}$$

For definite integrals not covered by the theorems of this section, evaluation by the method of residues, when possible at all, usually requires considerable ingenuity in selecting the appropriate contour and in eliminating the integrals over all but the desired portion of the contour. Several examples of this sort will be found, with hints, in the exercises.

EXERCISES

Evaluate the following integrals by the method of residues:

1 $\displaystyle\int_0^{2\pi}\frac{d\theta}{1-2p\sin\theta+p^2}\qquad -1<p<1$

2 $\displaystyle\int_0^{2\pi}\frac{d\theta}{(a+b\cos\theta)^2}\qquad 0<b<a$

3 $\displaystyle\int_0^{2\pi}\frac{d\theta}{\cos\theta+2\sin\theta+3}$

4 $\displaystyle\int_0^{2\pi}\frac{d\theta}{2\cos\theta+3\sin\theta+7}$

5 $\displaystyle\int_0^{2\pi}\frac{\sin^2\theta\,d\theta}{a+b\cos\theta}\qquad 0<b<a$

6 $\displaystyle\int_0^{\pi}\frac{\cos 2\theta\,d\theta}{5+4\cos\theta}$

7 $\displaystyle\int_{-\infty}^{\infty}\frac{dx}{x^4+a^4}$

8 $\displaystyle\int_{-\infty}^{\infty}\frac{dx}{(1+x^2)^3}$

9 $\displaystyle\int_{-\infty}^{\infty}\frac{x^2\,dx}{1+x^6}$

10 $\displaystyle\int_{-\infty}^{\infty}\frac{x^2\,dx}{(1+x^4)^2}$

11 $\displaystyle\int_0^{\infty}\frac{dx}{(a^2+x^2)^2}$

12 $\displaystyle\int_0^{\infty}\frac{dx}{1+x^6}$

13 $\displaystyle\int_{-\infty}^{\infty}\frac{\cos mx}{(x-a)^2+b^2}\,dx$

14 $\displaystyle\int_{-\infty}^{\infty}\frac{\sin mx}{(x-a)^2+b^2}\,dx$

15 $\displaystyle\int_0^{\infty}\frac{\cos mx}{(a^2+x^2)^2}\,dx$

16 $\displaystyle\int_0^{\infty}\frac{\cos mx}{1+x^4}\,dx$

17 $\displaystyle\int_{-\infty}^{\infty}\frac{\cos mx}{(x^2+a^2)(x^2+b^2)}\,dx$

18 $\displaystyle\int_{-\infty}^{\infty}\frac{x\sin mx}{(x^2+a^2)(x^2+b^2)}\,dx$

19 $\displaystyle\int_{-\infty}^{\infty}\frac{x\sin mx}{1+x^4}\,dx$

20 $\displaystyle\int_0^{\infty}\frac{x^{a-1}}{(x+b)(x+c)}\,dx\qquad 0<a<2\qquad 0<b,c$

21 $\displaystyle\int_0^{\infty}\frac{x^{a-1}}{(x-b)^2+c^2}\,dx\qquad 0<a<2$

22 $\displaystyle\int_0^{\infty}\frac{x^{a-1}}{(x+b)(x+c)(x+d)}\,dx\qquad 0<a<3\qquad 0<b,c,d$

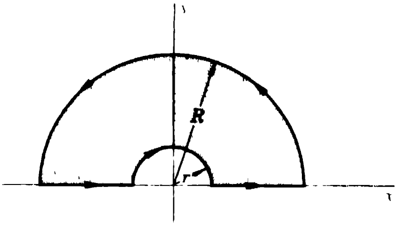
23 $\displaystyle\int_0^{\infty}\frac{x^{a-1}}{1+x^3}\,dx\qquad 0<a<3$

24 $\displaystyle\int_0^{\infty}\frac{x^{a-1}}{1+x^4}\,dx\qquad 0<a<4$

25 Show that $\Gamma(a)\Gamma(1-a) = \pi/(\sin a\pi)$ for $0 < a < 1$. [Hint. Consider the integral $\int_0^\infty \frac{y^{a-1}}{1+y}\,dy$, and evaluate it first by the method of residues and then by making the substitution $y = x/(1-x)$ and expressing it in terms of gamma functions.]

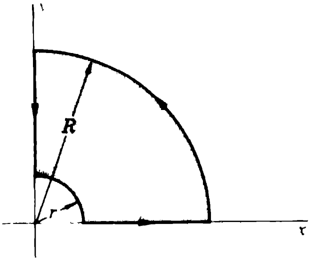
26 Show that $\int_0^\infty \frac{\sin x}{x}\,dx = \frac{\pi}{2}$. (Hint: Integrate e^{iz}/z around the contour shown in Fig 16.3, and let $r \to 0$ and $R \to \infty$.)

FIGURE 16.3



27 Show that $\int_0^\infty \frac{\cos x}{\sqrt{x}}\,dx = \int_0^\infty \frac{\sin x}{\sqrt{x}}\,dx = \sqrt{\frac{\pi}{2}}$. [Hint Integrate $e^{iz}/\sqrt{z}$ around the contour shown in Fig 16.4, let $r \to 0$ and $R \to \infty$, and recall (Exercise 10, Sec 7.3) that $\int_0^\infty e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}$.]

FIGURE 16.4

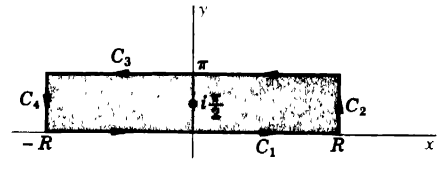


28 If $f(z)$ has a number of first-order poles on the real axis, but otherwise satisfies all the conditions of Theorem 2, show that the principal value of $\int_{-\infty}^{\infty} e^{imx}f(x)\,dx$ is equal to $2\pi i$ times the sum of the residues of $e^{imz}f(z)$ at its poles in the upper half plane plus $i\pi$ times the sum of the residues of $e^{imz}f(z)$ at its poles on the real axis. [Hint Use a contour like that shown in Fig 16.3, suitably indented around each of the poles of $f(z)$ which lies on the real axis.]

29 What is the Fourier expansion of the periodic function $\frac{1}{a + b\cos\theta}$ $(0 < b < a)$? Discuss from the point of view of Theorem 3, Sec. 6.3, the limiting behavior of the Fourier coefficients of this function as $n \to \infty$.

30 Show that $\int_{-\infty}^{\infty} \frac{\cos mx}{e^x + e^{-x}}\,dx = \frac{\pi}{e^{m\pi/2} + e^{-m\pi/2}}$. [Hint Integrate the function $e^{imz}/(e^z + e^{-z})$ around the contour shown in Fig. 16.5 and let $R \to \infty$.]

FIGURE 16.5



16.3

The complex inversion integral

We are now in a position to appreciate more fully the significance of the complex inversion integral of Laplace transform theory. In Sec. 6.8 we defined the Laplace transform of a function $f(t)$ to be

(1) $$\mathcal{L}\{f(t)\} = \int_0^\infty f(t)e^{-st}\,dt$$

and we showed that conversely

(2) $$f(t) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \mathcal{L}\{f(t)\}e^{st}\,ds$$

s being a complex variable. It is interesting now to reconsider the derivation of (2) in the light of complex variable theory and to investigate how this formula can be applied to the determination of a function when its transform is known.

In the complex plane let $\phi(z)$ be a function of z, analytic on the line $x = a$ and in the entire half plane R to the right of this line. Moreover, let $|\phi(z)|$ approach zero uniformly as z becomes infinite through this half plane. Then, if s is any point in the half plane R, we can choose a semicircular contour $C = C_1 + C_2$, as shown in Fig. 16.6, and apply Cauchy's integral formula, getting

(3) $$\phi(s) = \frac{1}{2\pi i}\int_C \frac{\phi(z)}{z-s}\,dz = \frac{1}{2\pi i}\int_{a+ib}^{a-ib}\frac{\phi(z)}{z-s}\,dz + \frac{1}{2\pi i}\int_{C_2}\frac{\phi(z)}{z-s}\,dz$$

Now, for values of z on the semicircle C_2 and for b sufficiently large, we have

$$|z-s| \geqq b - |s-a| \geqq b - |s| - |a|$$

whether a is positive, as shown in Fig. 16.6, or negative. Hence, letting M denote the maximum value of $|\phi(z)|$ on C_2, we have

$$\left|\int_{C_2}\frac{\phi(z)}{z-s}\,dz\right| \leqq \int_{C_2}\frac{|\phi(z)|}{|z-s|}\,|dz| \leqq \frac{M}{b-|s|-|a|}\int_{C_2}|dz| = \frac{\pi b M}{b-|s|-|a|}$$

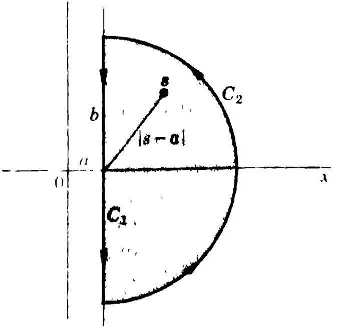


FIGURE 16.6
The contour used to obtain the complex inversion integral.

As b becomes infinite, the fraction

$$\frac{b}{b - |s| - |a|}$$

approaches 1, and at the same time M approaches zero, since, by hypothesis, $|\phi(z)|$ converges uniformly to zero as z becomes infinite through the right half plane R. Hence,

$$\lim_{b\to\infty} \int_{C_1} \frac{\phi(z)}{z - s}\, dz = 0$$

and in the limit we have, from (3),

$$\phi(s) = \lim_{b\to\infty} \frac{1}{2\pi i} \int_{a+ib}^{a-ib} \frac{\phi(z)}{z - s}\, dz = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} \frac{\phi(z)}{s - z}\, dz$$

Let us now attempt to determine the function of t whose Laplace transform is $\phi(s)$. Taking the inverse of $\phi(s)$ as defined by the last expression, we have

$$\mathcal{L}^{-1}\{\phi(s)\} = f(t) = \mathcal{L}^{-1}\left\{\frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} \frac{\phi(z)}{s - z}\, dz\right\}$$

Assuming that the operations of integrating along the vertical line $x = a$ and applying the inverse Laplace transformation can be interchanged, the last equation can be written

$$f(t) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} \mathcal{L}^{-1}\left\{\frac{\phi(z)}{s - z}\right\} dz$$

or, since the operator $\mathcal{L}^{-1}$ refers only to the variable s,

$$f(t) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} \phi(z)\mathcal{L}^{-1}\left\{\frac{1}{s - z}\right\} dz$$

Now the specific result

$$\mathcal{L}^{-1}\left\{\frac{1}{s - z}\right\} = e^{zt}$$

is known to us through independent reasoning Hence, we have finally

$$f(t) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} \phi(z)e^{zt}\, dz$$

which, except that the variable of integration is z instead of s, is exactly Eq. (2). From this result it is clear that *the inversion integral is a line integral in the complex plane, taken along a vertical line to the right of all the singularities of the transform $\phi(s)$ or along any other path into which this can legitimately be deformed.*

In the usual applications, the evaluation of the complex inversion integral is accomplished by the method of residues, using a semicircular contour whose diameter is the segment joining the points $a - ib$ and $a + ib$ and whose radius b is large enough to ensure that all the poles of the transform are within the contour (Fig. 16.7). Specifically, we have the following result

THEOREM 1

If $\phi(s)$ is an analytic function of s except at a finite number of poles each of which lies to the left of the vertical line $\Re(s) = a$ and if $s\phi(s)$ is bounded as s becomes infinite through the half plane $\Re(s) \leqq a$, then

$$\mathcal{L}^{-1}\{\phi(s)\} = \Sigma \text{ residues of } \phi(s)e^{st} \text{ at each of its poles}$$

PROOF Using the contour shown in Fig 16 7, we have, by the residue theorem,

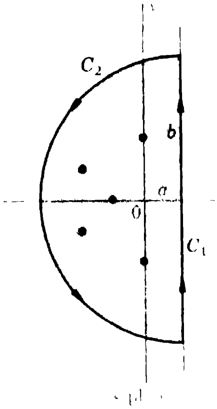


FIGURE 16.7
A typical contour used in the evaluation of the complex inversion integral

$$\frac{1}{2\pi i}\int_{a-ib}^{a+ib} \phi(s)e^{st}\,ds + \frac{1}{2\pi i}\int_{C_2} \phi(s)e^{st}\,ds = \sum \text{residues of } \phi(s)e^{st}$$

Hence,

$$\left|\frac{1}{2\pi i}\int_{a-ib}^{a+ib} \phi(s)e^{st}\,ds - \sum \text{residues of } \phi(s)\right| = \left|-\frac{1}{2\pi i}\int_{C_2} \phi(s)e^{st}\,ds\right| \tag{4}$$

Now along C_2 we have

$$s = a + be^{i\theta} \qquad \frac{\pi}{2} \leqq \theta \leqq \frac{3\pi}{2}$$

and, for sufficiently large s,

$$|s\phi(s)| < M \qquad \text{and} \qquad |s - a| \leqq |s| + |a| < 2|s|$$

Therefore,

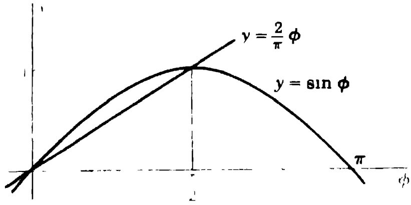
$$\begin{aligned}\left|-\frac{1}{2\pi i}\int_{C_2} \phi(s)e^{st}\,ds\right| &\leqq \frac{1}{2\pi}\int_{C_2} |\phi(s)|\,|e^{st}|\,|ds| \\ &= \frac{1}{2\pi}\int_{\pi/2}^{3\pi/2} |\phi(s)|\,\left|e^{t[a+b(\cos\theta+i\sin\theta)]}\right|\,|be^{i\theta}\,d\theta| \\ &\leqq \frac{1}{2\pi}\int_{\pi/2}^{3\pi/2} |\phi(s)|\,|s-a|e^{t(a+b\cos\theta)}\,d\theta \\ &\leqq \frac{1}{2\pi}\int_{\pi/2}^{3\pi/2} 2|s\phi(s)|e^{at}e^{bt\cos\theta}\,d\theta \\ &\leqq \frac{1}{2\pi}\,2Me^{at}\int_{\pi/2}^{3\pi/2} e^{bt\cos\theta}\,d\theta\end{aligned}$$

If we now set $\theta = \pi/2 + \alpha$ and then take advantage of the symmetry of the resulting integrand, the last integral becomes

$$\frac{M}{\pi} e^{at} \int_0^{\pi} e^{-bt\sin\alpha}\, d\alpha = \frac{2M}{\pi} e^{at} \int_0^{\pi/2} e^{-bt\sin\alpha}\, d\alpha$$

Now it is evident from Fig. 16.8 that

FIGURE 16.8
Plot showing that $\sin\phi \geqq 2\phi/\pi \geqq 0$ for $0 \leqq \phi \leqq \pi/2$



$$\sin\alpha \geqq \frac{2\alpha}{\pi} \qquad \text{for } 0 \leqq \alpha \leqq \frac{\pi}{2}$$

Hence the last integral is overestimated if we replace $\sin\alpha$ in the exponent by the smaller *positive* quantity $2\alpha/\pi$ Doing this and then performing the integration, we have

$$\left|\frac{1}{2\pi i}\int_{C_2} \phi(s)e^{st}\, ds\right| \leqq \frac{2M}{\pi} e^{at}\left[\frac{e^{-2bt\alpha/\pi}}{-2bt/\pi}\right]_0^{\pi/2}$$
$$= \frac{2M}{\pi} e^{at}\left[-\frac{\pi}{2bt}(e^{-bt} - 1)\right]$$

For $t \geqq 0$, the last expression clearly approaches zero as b becomes infinite Hence, returning to Eq. (4), it is clear that

$$\lim_{b\to\infty} \frac{1}{2\pi i}\int_{a-ib}^{a+ib} \phi(s)e^{st}\, ds \equiv \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \phi(s)e^{st}\, ds$$
$$\equiv \mathcal{L}^{-1}\{\phi(s)\}$$
$$= \Sigma \text{ residues of } \phi(s)e^{st} \qquad \text{as asserted}$$

The proof of the last theorem breaks down if $\phi(s)$ has infinitely many poles, because then, as $b \to \infty$, there will always be semicircles C_2 on which $|s\phi(s)|$ is not bounded However, by choosing a sequence of semicircles whose radii become infinite and no one of which passes through a pole of $\phi(s)$, it is possible to show that the result of the last theorem is still valid in the case when $\phi(s)$ has an infinite number of poles.*

EXAMPLE 1

What is $\mathcal{L}^{-1}\left\{\dfrac{1}{(s+a)^2 + b^2}\right\}$?

Using Theorem 1, we have only to compute the residues of

$$\frac{e^{st}}{(s+a)^2 + b^2}$$

---

* For a more detailed discussion of this point see, for instance, R V Churchill, "Operational Mathematics," 2d ed., pp 190–193, M Graw-Hill Book Company, New York, 1958.

at its two first-order poles $-a \pm ib$ At $s = -a + ib$, we have for the residue

$$\lim_{s \to -a+ib} \frac{[s - (-a + ib)]e^{st}}{[s - (-a + ib)][s - (-a - ib)]} = \frac{e^{(-a+ib)t}}{2ib}$$

and, at $s = -a - ib$, we have for the residue

$$\lim_{s \to -a-ib} \frac{[s - (-a - ib)]e^{st}}{[s - (-a + ib)][s - (-a - ib)]} = \frac{e^{(-a-ib)t}}{-2ib}$$

Hence, by Theorem 1,

$$\begin{aligned} f(t) = \mathcal{L}^{-1}\{\phi(s)\} &= \frac{e^{(-a+ib)t}}{2ib} + \frac{e^{(-a-ib)t}}{-2ib} \\ &= e^{-at}\frac{e^{ibt} - e^{-ibt}}{2ib} \\ &= \frac{e^{-at} \sin bt}{b} \end{aligned}$$

This example, of course, has been merely a new approach to a result with which we were already familiar. However, in more difficult applications the use of the complex inversion integral and contour integration is often either the only or, at least, the best way of finding a function when its transform is known

EXAMPLE 2

What is $\mathcal{L}^{-1}\left[\dfrac{1}{s \cosh as}\right]$?

Obviously in this case the function $\phi(s)$ has a first-order pole at $s = 0$ Moreover, since $\cosh as = \cos ias$ [Eq (13), Sec. 14 7], it follows that $\phi(s)$ has infinitely many other first-order poles, namely, the points where

$$ias = \pm\frac{(2n - 1)\pi}{2} \qquad \text{or} \qquad s = \pm\frac{(2n - 1)\pi}{2ia} \qquad n = 1, 2, 3,$$

However, if we set $s = \sigma + i\omega$, we have, by Eq (17), Sec 14.7,

$$\begin{aligned} |s\phi(s)| = \left|\frac{1}{\cosh as}\right| &= \frac{1}{\cosh^2 a\sigma \cos^2 a\omega + \sinh^2 a\sigma \sin^2 a\omega} \\ &= \frac{1}{\cosh^2 a\sigma - \sin^2 a\omega} \end{aligned}$$

and this is bounded on any semicircle which does not pass through one of the poles of $\phi(s)$. Hence, the inverse of $\phi(s)$ is simply the sum of the residues of

$$\frac{e^{st}}{s \cosh as}$$

at each of its poles, i e., the poles of $\phi(s)$

At $s = 0$ the residue is

$$\lim_{s \to 0} \frac{e^{st}}{\cosh as} = 1$$

and, at $s = \dfrac{(2n - 1)\pi}{2ia}$ [using l'Hospital's rule and Eq (20), Sec. 14 7, to evaluate the indeterminacy], the residue is

$$\lim_{s\to\frac{(2n-1)\pi}{2ia}} \frac{[s-(2n-1)\pi/2ia]e^{st}}{s\cosh as} = \frac{e^{(2n-1)\pi t/2ia}}{\dfrac{(2n-1)\pi}{2ia}\,a\sinh\dfrac{(2n-1)\pi}{2i}}$$

$$= \frac{2(-1)^n e^{(2n-1)\pi t/2ia}}{(2n-1)\pi}$$

Similarly, at $s = -\dfrac{(2n-1)\pi}{2ia}$, the residue is

$$\frac{2(-1)^n e^{-(2n-1)\pi t/2ia}}{(2n-1)\pi}$$

Hence, pairing the terms which correspond to the same value of n,

$$f(t) = \mathcal{L}^{-1}\{\phi(s)\} = 1 + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{2n-1}\left(e^{(2n-1)\pi t/2ia} + e^{-(2n-1)\pi t/2ia}\right)$$

$$= 1 + \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{2n-1}\cos\frac{(2n-1)\pi t}{2a}$$

EXERCISES

Using the complex inversion integral, find the inverses of the following Laplace transforms In each case discuss the resemblance of the method of residues to the use of the Heaviside expansion theorems (Sec 7 5)

1 $\dfrac{1}{(s+1)(s+3)}$

2 $\dfrac{1}{(s+2)^2}$

3 $\dfrac{1}{s^2+4}$

4 $\dfrac{s}{s^2+4s+13}$

5 $\dfrac{1}{s(s^2+1)}$

6 $\dfrac{s}{s^3+1}$

7 $\dfrac{s}{(s^2+4)^2}$

8 $\dfrac{1}{(s^2+9)(s^2+4)}$

9 $\dfrac{s+1}{(s+2)^2(s+3)}$

10 $\dfrac{1}{(s^2+2s+5)^2}$

11 Complete the solution of Exercise 8, Sec 8 7, by finding the angular displacement at a general point x

Find the inverse of each of the following transforms

12 $\dfrac{1}{s\sinh as}$

13 $\dfrac{1}{(s+b)\cosh as}$

14 $\dfrac{\sinh x\sqrt{s}}{s\sinh\sqrt{s}}$

15 $\dfrac{I_0(r\sqrt{s})}{sI_0(\sqrt{s})}$, where I_0 is the modified Bessel function of the first kind.

16.4

Stability criteria

In the analysis of many physical systems a complete description of the behavior of the system is unnecessar and all that is required is a knowledge of whether or not the system is stable, i.e., whether its response to a bounded excitation remains bounded or becomes infinite as $t \to \infty$. As we shall see in this section, this question can be answered by analyzing the Laplace transform of the response without actually determining the response itself.

We begin by supposing that, by methods such as those we described in Chap. 7, we have obtained the Laplace transform of the response of the system $\mathcal{L}\{y(t)\} \equiv \phi(s)$ and that $\phi(s)$ is a rational function; i e.,

$$\phi(s) = \frac{P(s)}{Q(s)}$$

where P and Q are real polynomials in the complex variable $s = a + i\omega$. Now we know from algebra that any polynomial, such as $Q(s)$, can always be factored into real linear and quadratic factors that may or may not be repeated Moreover, we know from the Heaviside theorems (Sec. 7 5) that the form of the inverse $y(t) = \mathcal{L}^{-1}\{\phi(s)\}$ is determined completely and solely by the factors of $Q(s)$ and that the only terms which can possibly occur in it are the following

Factor	Term
From unrepeated factors	
1 s	1
2 $s^2 + b^2$	$\cos bt$, $\sin bt$
3 $s - a$	e^{at}
4 $(s - a)^2 + b^2$	$e^{at} \cos bt$, $e^{at} \sin bt$
From repeated factors	
5 $s^n, n > 1$	$t^k, 0 < k \leqq n - 1$
6. $(s^2 + b^2)^n, n > 1$	$t^k \cos bt, t^k \sin bt, 0 < k \leqq n - 1$
7 $(s - a)^n, n > 1$	$t^k e^{at}, 0 < k \leqq n - 1$
8 $[(s - a)^2 + b^2]^n, n > 1$	$t^k e^{at} \cos bt, t^k e^{at} \sin bt, 0 < k \leqq n - 1$

Clearly, terms of the forms 1 and 2 are stable in all cases, for, although they do not approach zero as $t \to \infty$, they do remain finite. Terms of the forms 3, 4, 7, and 8 are stable if and only if a is negative, in which case they not only remain finite but in fact approach zero as $t \to \infty$ Terms of the forms 5 and 6 are unstable in all cases, since, because of the factor t, each becomes unbounded as $t \to \infty$. Translating these observations into conditions on the roots of the polynomial equation $Q(s) = 0$, we see that the response $y(t)$ will be stable if and only if the following conditions are met·

a Every unrepeated real root is nonpositive.
b Every repeated real root is negative.
c Every pure imaginary root is unrepeated.
d Every general complex root has negative real part.

Geometrically speaking, these conditions can be described as follows.

THEOREM 1

In order for the function

$$y(t) = \mathcal{L}^{-1}\left[\frac{P(s)}{Q(s)}\right]$$

to be stable, it is necessary and sufficient that the equation $Q(s) = 0$ have no roots to the right of the imaginary axis in the complex s-plane and that any root on the imaginary axis in the s-plane be unrepeated

Various methods are available for determining whether or not the roots of a polynomial equation all have nonpositive real parts * In general, however, these are more conveniently formulated as methods for determining whether or not the roots all have real parts that are strictly negative, and most, though not all, of our results will be of this nature This is not a serious disadvantage, because in practice zero roots and pure imaginary roots, i e , roots whose real parts are zero, if they occur at all, are usually easily recognizable

A preliminary result of considerable importance is contained in the following theorem.

THEOREM 2

The real part of each root of the polynomial equation $Q(s) = 0$ is less than or equal to zero only if the coefficients in $Q(s)$ all have the same sign

PROOF We observe first that it is no specialization to interpret the condition of the theorem as asserting that all coefficients in $Q(s)$ are positive For the case in which all coefficients are negative can be converted into the case in which all coefficients are positive, and vice versa, simply by multiplying $Q(s) = 0$ by -1, which, of course, in no way alters the roots of this equation Now, if every root of $Q(s) = 0$ has nonpositive real part, then the only possible factors of $Q(s)$ are of the forms

$$s + a_i \qquad \text{and} \qquad (s + a_j)^2 + b_j^2 \qquad \text{where } a_i,\ a_j \geqq 0$$

Since these factors contain only nonnegative terms and since $Q(s)$ is simply the product of a finite number of these factors, it is clear that every nonzero coefficient in $Q(s)$ must be positive, as asserted

It is also clear from the preceding argument that, if every a is positive, so that all roots of $Q(s) = 0$ have real parts strictly negative, then there can be no zero coefficients in $Q(s)$; i.e , all terms must be present Hence, restating this observation contrapositively, we have the following corollary

---

* See, for instance, A Bronwell, "Advanced Mathematics in Physics and Engineering," pp. 386–413, McGraw-Hill Book Company, New York, 1953, and E A. Guillemin, "The Mathematics of Circuit Analysis," pp. 395–409, John Wiley & Sons, Inc , New York, 1953

COROLLARY 1

If one or more terms are missing from $Q(s)$, then the equation $Q(s) = 0$ has at least one root whose real part is nonnegative.

The condition of Theorem 2 is only a necessary and not a sufficient one, that is, it *cannot* be asserted, conversely, that, if the coefficients in $Q(s)$ all have the same sign, then the real part of each root of $Q(s) = 0$ is nonpositive. For instance,

$$s^4 + s^3 + s^2 + 11s + 10$$

contains only terms with positive coefficients, yet the roots of the equation

$$s^4 + s^3 + s^2 + 11s + 10 = 0$$

are $\quad s = -1, -2, 1 \pm 2i$

and the two complex roots have positive real parts On the other hand, it is clear from Theorem 2 that we do have the following result

COROLLARY 2

If $Q(s)$ contains some terms with positive coefficients and some terms with negative coefficients, then the equation $Q(s) = 0$ has at least one root whose real part is positive

For quadratic equations the necessary condition of Theorem 2 is also sufficient. For if the equation $a_0s^2 + a_1s + a_2 = 0$ contains no negative coefficients, then its roots

$$s = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_0a_2}}{2a_0}$$

are clearly either nonpositive real numbers or conjugate complex numbers with nonpositive real parts

For cubic equations, a sufficient condition, supplementing Theorem 2, is contained in the following result

THEOREM 3

A necessary and sufficient condition that every root of the cubic equation $a_0s^3 + a_1s^2 + a_2s + a_3 = 0$ have negative real part is that all coefficients have the same sign and that $a_1a_2 - a_0a_3 > 0$

PROOF Let us assume for definiteness that the given equation has one real root r and one pair of conjugate complex roots $p \pm iq$. The case in which the equation has three real roots can be handled in exactly the same fashion From algebra we recall that the roots, say r_1, r_2, r_3, of any cubic equation are related to the coefficients through the equations

$$\frac{a_1}{a_0} = -(r_1 + r_2 + r_3)$$

$$\frac{a_2}{a_0} = r_1r_2 + r_2r_3 + r_3r_1$$

$$\frac{a_3}{a_0} = -r_1r_2r_3$$

In the present case these become

$$(1) \qquad \frac{a_1}{a_0} = -(r + 2p)$$

$$(2) \qquad \frac{a_2}{a_0} = p^2 + q^2 + 2pr$$

$$(3) \qquad \frac{a_3}{a_0} = -r(p^2 + q^2)$$

From (3) and the assumption that the a's all have the same sign, it follows that $r < 0$ To prove that $p < 0$, we note that the condition $a_1a_2 - a_0a_3 > 0$ can be rewritten, after division by a_0^2, as

$$\frac{a_1}{a_0}\frac{a_2}{a_0} - \frac{a_3}{a_0} > 0$$

When the ratios of the a's are replaced by their equivalents from (1), (2), and (3), this becomes

$$-(r + 2p)(p^2 + q^2 + 2pr) + r(p^2 + q^2) > 0$$

or, simplifying and rearranging,

$$(4) \qquad -2p[(p^2 + q^2 + 2pr) + r^2] > 0$$

Now from (2) and the hypothesis that the a's are all of the same sign, it is evident that $p^2 + q^2 + 2pr > 0$ Hence, $(p^2 + q^2 + 2pr) + r^2 > 0$, and it follows from (4) that $p < 0$, as asserted This proves the sufficiency of the conditions of Theorem 3.

The necessity that all the coefficients have the same sign follows immediately from (1), (2), and (3), since the right-hand sides of these relations are all positive if $p < 0$ and $r < 0$ The necessity of the condition $a_1a_2 - a_0a_3 > 0$ follows by reversing the above steps and working backward to this inequality from (4), which is surely true if $p < 0$ and $r < 0$

The extension of Theorem 3 to polynomial equations of higher degree is contained in the next theorem, which we state without proof.*

THEOREM 4

In the polynomial equation

$$Q(s) = a_0s^n + a_1s^{n-1} + a_2s^{n-2} + \qquad + a_{n-1}s + a_n = 0$$

let every coefficient be positive, and construct the n quantities

$$D_1 = a_1 \qquad D_2 = \begin{vmatrix} a_1 & a_0 \\ a_3 & a_2 \end{vmatrix} \qquad D_3 = \begin{vmatrix} a_1 & a_0 & 0 \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{vmatrix}$$

$$D_n = \begin{vmatrix} a_1 & a_0 & 0 & 0 & 0 & 0 & & & \\ a_3 & a_2 & a_1 & a_0 & 0 & 0 & & & \\ a_5 & a_4 & a_3 & a_2 & a_1 & a_0 & & & \\ \cdot\ \cdot & \cdot\ \cdot & \cdot & \cdot\ \cdot & & \cdot & & \cdot & \cdot \\ a_{2n-1} & a_{2n-2} & a_{2n-3} & a_{2n-4} & a_{2n-5} & a_{2n-6} & \cdot & \cdot & a_n \end{vmatrix}$$

---

* See, for instance, J V. Uspensky, "Theory of Equations," pp 304–309, McGraw-Hill Book Company, New York, 1948

where, in each determinant, all a's with negative subscripts or with subscripts greater than n are to be replaced by zero. Then a necessary and sufficient condition that each root of $Q(s) = 0$ have negative real part is that each D_n be positive.

This is commonly known as the **Routh** or **Routh-Hurwitz stability criterion.**

EXAMPLE 1

For the equation $s^5 + s^4 + 2s^3 + s^2 + s + 2 = 0$, we have

$$D_1 = 1 \qquad D_2 = \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} = 1 \qquad D_3 = \begin{vmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 2 & 1 & 1 \end{vmatrix} = 2$$

$$D_4 = \begin{vmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 1 \\ 2 & 1 & 1 & 2 \\ 0 & 0 & 2 & 1 \end{vmatrix} = -4 \qquad D_5 = \begin{vmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 \\ 2 & 1 & 1 & 2 & 1 \\ 0 & 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{vmatrix} = -8$$

Since not all of the D's are positive, the given equation has at least one root whose real part is nonnegative This can be confirmed, of course, by actually finding the roots of the given equation, which are, in fact

$$r_1 = -1 \qquad r_2, r_3 = \tfrac{1}{2} \pm i\sqrt{3}/2 \qquad r_4, r_5 = -\tfrac{1}{2} \pm i\sqrt{7}/2$$

A somewhat different method of obtaining information about the location of the roots of an equation $f(z) = 0$, which has the advantage that it tells exactly how many roots there are with positive real parts and, moreover, is not restricted to the case where $f(z)$ is a polynomial, is based on the following theorem.

THEOREM 5

If $f(z)$ is analytic within and on a closed curve C except at a finite number of poles and if $f(z)$ has neither poles nor zeros on C, then

$$\frac{1}{2\pi i}\int_C \frac{f'(z)}{f(z)}\,dz = N - P$$

where N is the number of zeros of $f(z)$ within C, and P is the number of poles of $f(z)$ within C, each counted as many times as its multiplicity.

PROOF Suppose first that, at a point $z = a_k$ within C, $f(z)$ has a zero of order n_k Then $f(z)$ can be written

$$f(z) = (z - a_k)^{n_k}\phi(z)$$

where $\phi(z)$ is nonvanishing and analytic in some neighborhood of $z = a_k$. From this

$$f'(z) = n_k(z - a_k)^{n_k-1}\phi(z) + (z - a_k)^{n_k}\phi'(z)$$

and thus

$$\frac{f'(z)}{f(z)} = \frac{n_k(z - a_k)^{n_k-1}\phi(z) + (z - a_k)^{n_k}\phi'(z)}{(z - a_k)^{n_k}\phi(z)} = \frac{n_k}{z - a_k} + \frac{\phi'(z)}{\phi(z)}$$

Since $\phi(z)$, and hence $\phi'(z)$, is analytic at $z = a_k$ and since $\phi(z)$ does not vanish at $z = a_k$, the fraction $\phi'(z)/\phi(z)$ is analytic at $z = a_k$ Hence it is clear from the last expression that $f'(z)/f(z)$ has a simple pole with residue n_k at every point a_k where $f(z)$ has a zero of order n_k. Similarly, if $f(z)$ has a pole of order p_k at the point $z = b_k$, we can write

$$f(z) = \frac{c_{-p_k}}{(z - b_k)^{p_k}} + \frac{c_{-p_k+1}}{(z - b_k)^{p_k-1}} + \cdot \quad \cdot + \frac{c_{-1}}{z - b_k} + c_0 + \cdot$$

Hence, putting these fractions over a common denominator, we have, in the neighborhood of $z = b_k$,

$$f(z) = \frac{1}{(z - b_k)^{p_k}} \psi(z) = (z - b_k)^{-p_k}\psi(z)$$

where $\psi(z) = c_{-p_k} + c_{-p_k+1}(z - b_k) + c_{-p_k+2}(z - b_k)^2 + \cdot$

is obviously analytic and nonvanishing at $z = b_k$. Therefore, around b_k,

$$f'(z) = -p_k(z - b_k)^{-p_k-1}\psi(z) + (z - b_k)^{-p_k}\psi'(z)$$

and thus $$\frac{f'(z)}{f(z)} = \frac{-p_k(z - b_k)^{-p_k-1}\psi(z) + (z - b_k)^{-p_k}\psi'(z)}{(z - b_k)^{-p_k}\psi(z)} = \frac{-p_k}{z - b_k} + \frac{\psi'(z)}{\psi(z)}$$

The last fraction on the right is clearly analytic; hence, $f'(z)/f(z)$ has a simple pole with residue $-p_k$ at every point where $f(z)$ has a pole of order p_k. Applying the residue theorem to $f'(z)/f(z)$ over the region bounded by C, we therefore have

$$\int_C \frac{f'(z)}{f(z)}\,dz = 2\pi i \sum \text{residues} = 2\pi i \left(\sum n_k - \sum p_k\right) = 2\pi i(N - P)$$

since Σn_k is the total multiplicity N of all the zeros of $f(z)$ within C and Σp_k is the total multiplicity P of all the poles of $f(z)$ within C. Dividing by $2\pi i$, we obtain the assertion of the theorem.

An important alternative form of the last theorem can be derived by noting that

$$\frac{1}{2\pi i}\int_C \frac{f'(z)}{f(z)}\,dz = \frac{1}{2\pi i}\int_C d[\ln f(z)]$$

Hence, performing the integration,

$$N - P = \frac{1}{2\pi i}[\text{variation of } \ln f(z) \equiv \ln|f(z)| + i \arg f(z) \text{ in going completely around } C]$$

Clearly, $\ln|f(z)|$ is the same at the beginning and at the end of any closed curve, and therefore

$$N - P = \frac{1}{2\pi i}[\text{variation of } i \arg f(z) \text{ around } C]$$

$$= \frac{\text{variation of } \arg f(z) \text{ around } C}{2\pi}$$

In particular, if $f(z)$ is analytic everywhere within C (so that $P = 0$), we have the following important result, commonly known as the **principle of the argument**:

COROLLARY 1

If $f(z)$ is analytic within and on a closed curve C and does not vanish on C, then the number of zeros of $f(z)$ within C is equal to $1/2\pi$ times the net variation in the argument of $f(z)$ as z traverses the curve C in the counterclockwise sense.

In geometric terms, this means that, if the locus of $w = f(z)$ is plotted for values of z ranging around the given contour C, then the number of times this locus encircles the origin in the w-plane is the number of zeros of $f(z)$ within C. Moreover, since $f(z) = 0$ implies $w = 0$, it is evident that if $f(z)$ has a zero on C, the image curve passes through the origin in the w-plane

To use the last theorem and its corollary to determine whether or not each of the roots of a polynomial equation $Q(z) = 0$ has negative real part, we proceed as follows In the z-plane let the contour C consist of the segment of the imaginary axis between $-R$ and R and the semicircle lying in the right half plane and having this segment as diameter (Fig. 16.9) Since a polynomial equation has only a finite number of roots, it is clear that, if R is taken sufficiently large, any roots of $Q(z) = 0$ which lie in the right half plane, i e , any roots which have positive real parts, will lie within C

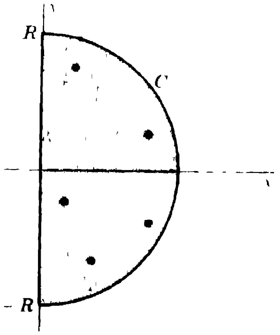


FIGURE 16 9
A semicircular contour enclosing all zeros of a function which lie in the right half plane

Now let z range over the contour C, and in an auxiliary w-plane let the locus of the corresponding values of $w = Q(z)$ be plotted If this curve does not enclose the origin in the w-plane, then according to the corollary of Theorem 5, $Q(z) = 0$ has no roots in the right half plane If, further, this curve does not pass through the origin in the w-plane, then $Q(z) = 0$ has no roots on the imaginary axis either, i e , all roots of $Q(z) = 0$ have negative real parts On the other hand, if the image curve encircles the origin in the w-plane a net number of times k, then $Q(z) = 0$ has k roots in the right half plane, i e , has k roots with positive real part Moreover, for every time this curve passes through the origin in the w-plane there is a root of $Q(z) = 0$ lying on the imaginary axis in the z-plane Distinct pure imaginary roots of $Q(z) = 0$ thus give rise to a multiple point at the origin in the w-plane, the tangents at the multiple point being distinct A repeated pure imaginary root in the z-plane similarly gives rise, in general, to a cusp at the origin in the w-plane

The labor of plotting the image curve in the w-plane can be reduced considerably by letting $R \rightarrow \infty$ The image of the semicircular portion of C then recedes to infinity in the w-plane, and without any plotting, its contribution to possible encirclements of the origin can be determined as follows. On the semicircle we have

$$z = Re^{i\theta} \qquad -\frac{\pi}{2} \leqq \theta \leqq \frac{\pi}{2}$$

For the images of these values of z we have

$$w = Q(Re^{i\theta}) = a_0(Re^{i\theta})^n + a_1(Re^{i\theta})^{n-1} + \cdots + a_n$$

Now, for arbitrarily large values of R, all terms in $Q(Re^{i\theta})$ after the first are negligible in comparison with the first term $a_0R^ne^{in\theta}$ Hence, as z traverses the semicircular portion of C in the positive direction, with $\theta = \arg z$ varying from $-\pi/2$ to $\pi/2$, the argument of its image

$$w \doteq a_0R^ne^{in\theta}$$

varies from $-n\pi/2$ to $n\pi/2$, which represents a net variation in $\arg w$, that is, $\arg Q(z)$, of $n\pi$ Hence if $w = Q(z)$ is plotted only for z varying from $i\infty$ to $-i\infty$ along the imaginary axis and the net change in the argument of w is noted, with its proper sign, of course, this change plus $n\pi$ will give the net change as the entire contour C is traversed. This change, divided by 2π, gives the net number of times the image curve encircles the origin in the w-plane, and this number is equal to the number of roots of $Q(z) = 0$ in the right half of the z-plane The labor of plotting can be still further reduced by noting that, for polynomials with real coefficients, such as we encounter in Laplace transforms, we have

$$Q(\bar{z}) = \overline{Q(z)}$$

and, hence, the plot of $Q(z)$ for values on the lower half of the imaginary axis is just the reflection in the real axis of the plot of $Q(z)$ for values of z on the upper half of the imaginary axis

EXAMPLE 2

Discuss the stability of $y(t)$ if $\mathcal{L}\{y(t)\} = (s^2 + 1)/(s^3 + s^2 + 4s + 1)$

As we pointed out above, the stability of $y(t)$ is determined solely by the location of the zeros of the denominator of $y(t)$ Hence, we begin by plotting

$$w = Q(s) = s^3 + s^2 + 4s + 1$$

for values of s on the imaginary axis, i e , for $s = i\omega$ and ω ranging from ∞ to $-\infty$ The parametric equations of the image curve are easily obtained, for

$$Q(i\omega) = -i\omega^3 - \omega^2 + 4i\omega + 1$$

and so the real and imaginary parts of $w = u + iv$ are

$$u = 1 - \omega^2 \qquad \text{and} \qquad v = 4\omega - \omega^3$$

Figure 16 10 shows a plot of this curve together with a plot of $\arg w$ Evidently, as s traverses the imaginary axis from $i\infty$ to $-i\infty$, $\arg w$ varies from $3\pi/2$ to $-3\pi/2$, which is a net variation of -3π This, added to the value $n\pi = 3\pi$ contributed by the semicircular portion of the contour C

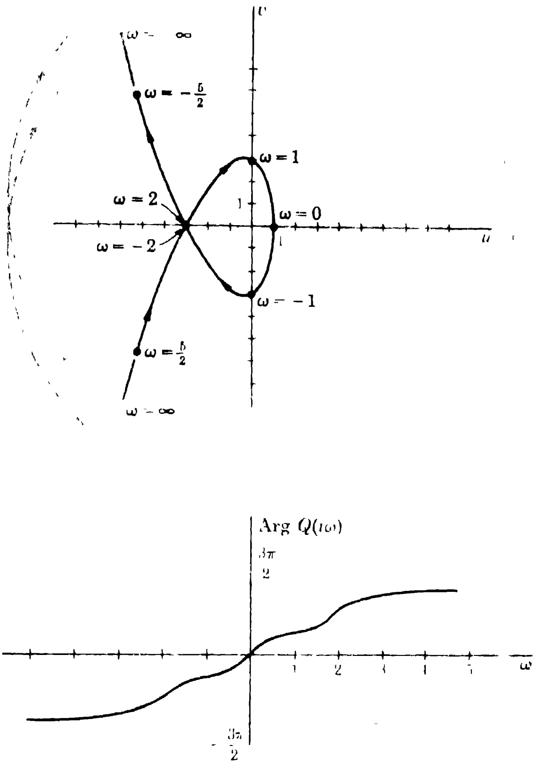


FIGURE 16.10
Plots of $Q(s) = s^3 + s^2 + 4s + 1$ and arg $Q(s)$ for $s = i\omega$

(Fig 16 9), gives a net variation of zero as the entire contour C is traversed. Hence, $Q(s)$ has no zeros in the right half of the s-plane. Moreover, since the image curve does not pass through the origin in the w-plane, $Q(s)$ has no zeros on the imaginary axis Therefore, by our earlier discussion, the inverse $y(t)$ is stable

EXAMPLE 3

Discuss the stability of $y(t)$ if $\mathcal{L}\{y(t)\} = (s-2)/(s^3+s^2+s+4)$.

Proceeding exactly as in Example 2, we obtain from

$$Q(i\omega) = -i\omega^3 - \omega^2 + i\omega + 4$$

the parametric equations

$$u = 4 - \omega^2 \quad \text{and} \quad v = \omega - \omega^3$$

and the image curve shown in Fig 16.11 In this case, as s traverses the imaginary axis from $i\infty$ to $-i\infty$, arg w varies from $3\pi/2$, as in Example 2, to $5\pi/2$, which is a net variation of $5\pi/2 - 3\pi/2 = \pi$ Hence, adding the variation $n\pi = 3\pi$ contributed by the semicircular portion of the contour C (Fig 16 9), we obtain 4π for the net variation in arg w as the entire contour C is traversed. Dividing this by 2π, we obtain 2 as the number of zeros of $Q(s)$ in the right half plane. The inverse in this case is, therefore, unstable

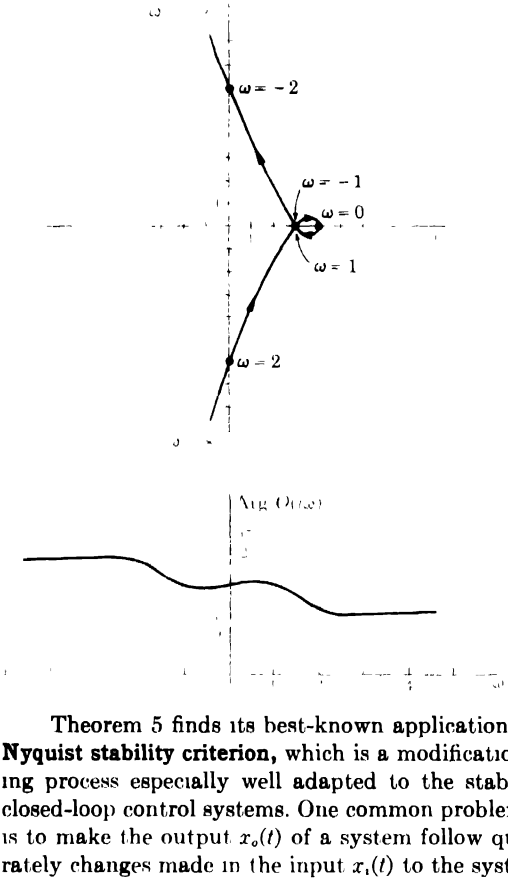


FIGURE 16.11
Plots of $Q(s) = s^3 + s^2 + s + 4$ and arg $Q(s)$ for $s = i\omega$

Theorem 5 finds its best-known application in the so-called **Nyquist stability criterion,** which is a modification of the preceding process especially well adapted to the stability analysis of closed-loop control systems. One common problem in engineering is to make the output $x_o(t)$ of a system follow quickly and accurately changes made in the input $x_i(t)$ to the system. In an **open-loop system,** such as that shown in Fig 16.12*a*, this is often difficult to accomplish; specifically, prolonged oscillation of $x_o(t)$ about its desired value may well follow an abrupt change of the input $x_i(t)$ to some desired new value One possible way to remedy this situation is to construct a **feedback loop,** such as the one shown in Fig 16.12*b*, which will sample the output and feed it

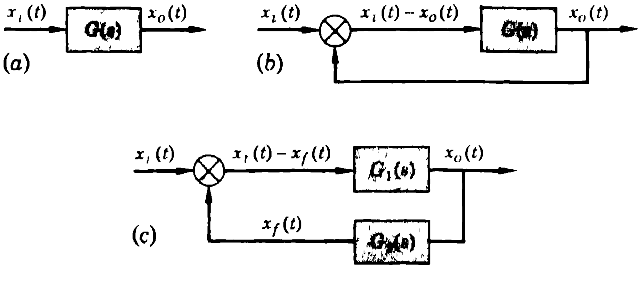


FIGURE 16.12
Systems with feedback loops

back to a differential device which will in turn transmit the **error signal** $x_i(t) - x_o(t)$ as a modified or corrected input to the original system More generally, the output $x_o(t)$ may be and usually is modified by some additional device in the feedback loop to produce the **feedback signal** $x_f(t)$ before it is fed to the differential (Fig. 16.12c)

In Fig 16 12c, let $G_1(s)$ and $G_2(s)$ be the transfer functions of the original system and the feedback loop, respectively. Then, from the definition of a transfer function as the ratio of the transformed output to the transformed input (Sec 7.7), we can write

$$\mathcal{L}\{x_o(t)\} = G_1(s)[\mathcal{L}\{x_i(t)\} - \mathcal{L}\{x_f(t)\}]$$

$$\mathcal{L}\{x_f(t)\} = G_2(s)\mathcal{L}\{x_o(t)\}$$

If we eliminate $\mathcal{L}\{x_f(t)\}$ between these two equations, we obtain at once

$$\mathcal{L}\{x_o(t)\} = \frac{G_1(s)}{1 + G_1(s)G_2(s)}\mathcal{L}\{x_i(t)\}$$

Evidently $G_1(s)/[1 + G_1(s)G_2(s)]$ is the over-all transfer function of the entire closed-loop system

The question of the stability of a feedback system is of great importance and, as we discussed above, can be answered by an examination of the Laplace transform of the output, namely,

$$\frac{G_1(s)}{1 + G_1(s)G_2(s)}\mathcal{L}\{x_i(t)\}$$

Now, if the original system without the feedback loop is stable for the input $x_i(t)$, as we shall suppose, then the product $G_1(s)\mathcal{L}\{x_i(t)\}$ can have no poles in the right half of the s-plane, and the stability of the over-all system depends solely on the location of the zeros of the denominator,

$$1 + G_1(s)G_2(s)$$

Hence, as before, we plot the locus of the function

$$w(s) = 1 + G_1(s)G_2(s)$$

as s ranges over the contour of Fig 16 9.

In this case, since $G_1(s)$ and $G_2(s)$ are themselves Laplace transforms, each approaches zero as R becomes infinite (Corollary 1, Theorem 5, Sec. 7 1). Hence, the image of the semicircular portion of the contour C shrinks to the single point $w = 1$ as $R \to \infty$. Thus, to determine stability, it is necessary only to plot $w(s) = 1 + G_1(s)G_2(s)$ for values of s on the imaginary axis and determine whether or not the resulting curve encloses the origin. Moreover, as we pointed out above, this curve can be constructed simply by plotting $1 + G_1(i\omega)G_2(i\omega)$ for positive values of ω and then reflecting the resulting arc in the real axis. In practice, instead of plotting $w = 1 + G_1(i\omega)G_2(i\omega)$ and observing whether or not the image curve encircles the origin, it is customary to plot $w = G_1(i\omega)G_2(i\omega)$ and observe whether or not it encircles the point $w = -1$. The equivalence of these two procedures is obvious.

It would take us too far afield and involve us in too many details of a purely engineering nature to discuss the application of the Nyquist stability criterion to specific, nontrivial closed-loop systems. Such applications appear in large numbers in books on servomechanisms, and to these we must refer for illustrations and further information.*

EXERCISES

Using the geometric approach based on the corollary of Theorem 5, determine whether or not the following equations have any roots with nonnegative real parts Check by using Theorem 4

1 $s^3 + s + 9 = 0$

2 $s^3 + 6s^2 + 10s + 6 = 0$

3 $s^4 + 2s^3 + 7s^2 + 4s + 10 = 0$

4 $s^4 + s^3 + s^2 + 10s + 10 = 0$

5 Prove Theorem 3 on the assumption that the cubic has three real roots

---

* See, for instance, G J Thaler and R G Brown, "Analysis and Design of Feedback Control Systems," 2d ed , McGraw-Hill Book Company, New York, 1960, or H Chestnut and R W Mayer, "Servomechanisms and Regulating System Design," John Wiley & Sons, Inc , New York, 1951

CHAPTER SEVENTEEN

Conformal Mapping

17.1

The geometrical representation of functions of z

Although in the last section we plotted the values of a function $w = f(z)$ for *certain* values of z, namely, those on a particular semicircular contour, we have not as yet attempted to provide a geometrical representation for $w = f(z)$ when z ranges over the *entire* complex plane. To do so now requires a decided departure from the conventional methods of cartesian plotting, which associate a curve with a real function $y = g(x)$ and a surface with a real function $z = h(x,y)$. In the complex domain, a functional relation $w = f(z)$, that is,

$$u + iv = f(x + iy)$$

involves *four* real variables, namely, the two independent variables x and y and the two dependent variables u and v. Hence, a space of *four* dimensions is required if we are to plot $w = f(z)$ in the cartesian fashion. To avoid the difficulties inherent in such a device, we choose instead to proceed as follows:

Let there be given two planes, one the z-plane, in which the point $z = x + iy$ is to be plotted, and the other the w-plane, in which the point $u + iv$ is to be plotted. A function $w = f(z)$ is now represented not by a locus of points in a space of four dimensions but by a correspondence between the points of these two cartesian planes. Whenever a point is given in the z-plane, the function $w = f(z)$ determines one or more values of $u + iv$ and, hence, one or more points in the w-plane. As z ranges over any configuration in the z-plane, the corresponding point $u + iv$ describes some configuration in the w-plane. The function $w = f(z)$ thus defines a **mapping** or a **transformation** of the z-plane onto the w-plane and, in turn, is represented geometrically by this mapping.

EXAMPLE 1

Discuss the way in which the z-plane is mapped onto the w-plane by the function $w = z^2$

In this case we have $u + iv = (x + iy)^2 = (x^2 - y^2) + 2ixy$, and thus

$$u = x^2 - y^2 \qquad v = 2xy \tag{1}$$

These are the equations of the transformation between the two planes From them, many features of the correspondence can easily be inferred

For instance, lines parallel to the y-axis, i e , lines with equations $x = c_1$, map into curves in the w-plane whose parametric equations are, from (1),

$$u = c_1^2 - y^2 \qquad v = 2c_1y$$

Eliminating the parameter y, we obtain the equation

$$u = c_1^2 - \frac{v^2}{4c_1^2}$$

This defines a family of parabolas having the origin of the w-plane as focus, the line $v = 0$ as axis, and all opening to the left (Fig 17 1) Similarly, lines parallel to the x-axis, i e , lines with

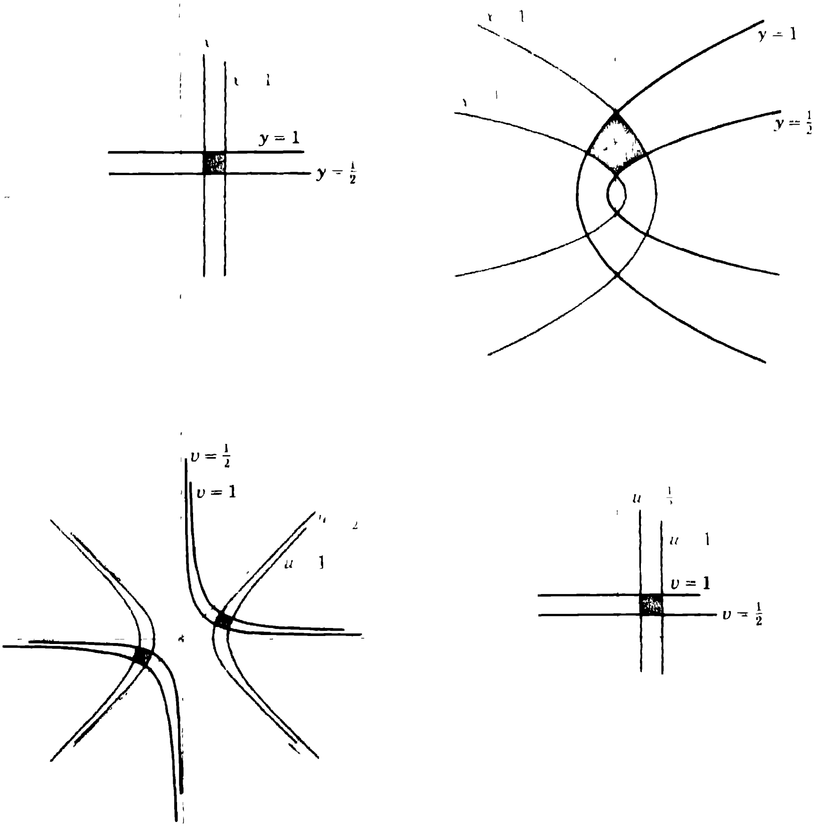


FIGURE 17.1

Plot showing the mapping of certain lines by the function $w = z^2$

equations $y = c_2$, map into curves in the w-plane whose parametric equations are

$$u = x^2 - c_2^2 \qquad v = 2c_2x$$

Eliminating x, we obtain

$$u = \frac{v^2}{4c_2^2} - c_2^2$$

which is the equation of a family of parabolas having the origin as focus, the line $v = 0$ as axis, but this time all opening to the right

Mapping from the w-plane back onto the z-plane is even more immediate From (1), the lines $u = k_1$ correspond to the rectangular hyperbolas

$$x^2 - y^2 = k_1$$

The lines $v = k_2$ correspond to the rectangular hyperbolas

$$xy = \tfrac{1}{2}k_2$$

The images of other curves, or regions, can, with varying degrees of difficulty, be found in the same fashion For instance, to find the curve into which the line

$$y = 2x + 1$$

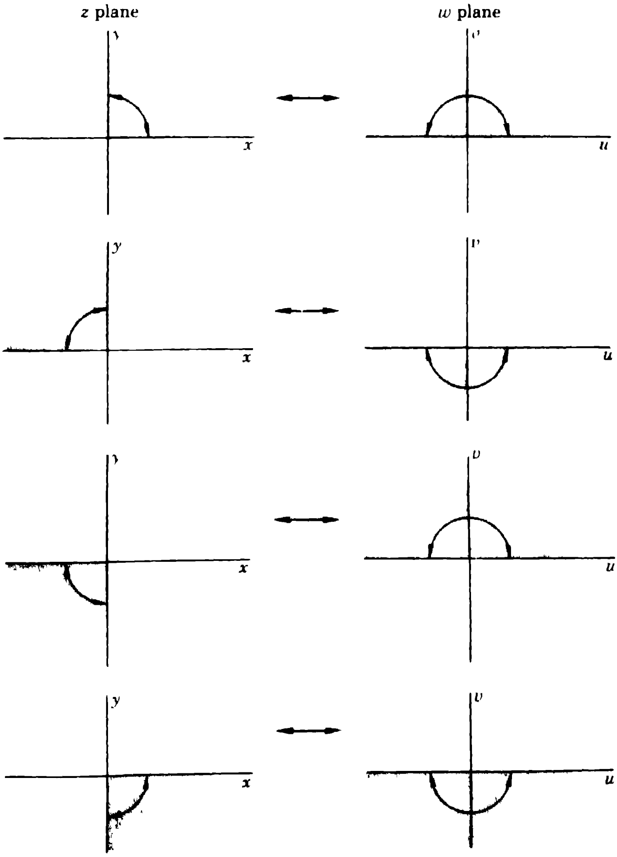


FIGURE 17.2 Plot illustrating the two-valued character of the mapping defined by $z = w^{1/2}$

is transformed, we must eliminate x and y between this equation and the equations of the transformation. To do this, we first substitute for y in Eqs. (1), getting

$$u = x^2 - (2x+1)^2 = -3x^2 - 4x - 1$$

$$v = 2x(2x+1) = 4x^2 + 2x$$

Solving these equations for x and x^2, we find at once

$$x = \frac{4u+3v+4}{-10} \qquad x^2 = \frac{u+2v+1}{5}$$

Hence,

$$\frac{u+2v+1}{5} = \left(\frac{4u+3v+4}{-10}\right)^2$$

or

$$16u^2 + 24uv + 9v^2 + 12u - 16v = 4$$

which is the equation of a parabola.

Although w is a single-valued function of z, the converse is not true. In fact, when w is given, z may be either of the two square roots of w. Because of this, the mapping from the z-plane to the w-plane covers the latter twice, as Fig. 17.2 shows. This, of course, is nothing but a graphic representation of the now familiar fact that the angles of complex numbers are doubled when the numbers are squared.

EXERCISES

1. Discuss the mapping between the z- and w-planes defined by the function $w = (\bar{z})^2$.
2. Discuss the transformation between the z- and w-planes defined by $w = x - iy$.
3. What relation, if any, exists between the transformations $w = f(z)$ and $w = f(\bar{z})$?
4. Discuss the transformation defined by $w = 2iz + 1$.
5. Discuss the transformation defined by $w = (x^2 - y^2) + ixy$. In what significant way does it differ from the transformation defined by $w = z^2 = (x^2 - y^2) + 2ixy$?
6. Discuss the transformation defined by $w = z^3$. Plot the image of the line $u = 1$. What is the equation of the image of the line $x = 1$?
7. Discuss the transformation defined by $w = z^4$. Plot the image of the line $u = 1$. What is the equation of the image of the line $x = 1$?
8. Discuss the transformation defined by the function $w = 1/z$. Plot the image of the square whose vertices are the points $z = 1 + i,\ 2 + i,\ 2 + 2i,\ 1 + 2i$.
9. Find the equations of the transformation defined by the function $(z - i)/z$, and show that every circle through the origin in the z-plane is transformed into a straight line in the w-plane.
10. Discuss the transformation defined by $w = e^z$. What is the equation of the image of the line $x + y = 1$?

17.2

Conformal mapping

In the last section we saw that every function of a complex variable maps the xy-plane onto the uv-plane. We now propose to investigate in more general terms the character of this transformation when the mapping function $w = u(x,y) + iv(x,y)$ is analytic.

At the outset it is important to know when the transformation equations can be solved (at least theoretically) for x and y as single-valued functions of u and v; that is, when the transformation has a single-valued inverse. The condition for this, as established in most texts on advanced calculus,* is simply that the Jacobian determinant of the transformation,

$$J\left(\frac{u,v}{x,y}\right) = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix}$$

be different from zero. Since $w = f(z)$ is assumed to be analytic, u and v must satisfy the Cauchy-Riemann equations. Hence, substituting into the Jacobian, we have

$$J\left(\frac{u,v}{x,y}\right) = \begin{vmatrix} \dfrac{\partial u}{\partial x} & -\dfrac{\partial v}{\partial x} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial u}{\partial x} \end{vmatrix} = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 = \left|\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right|^2 = |f'(z)|^2$$

which establishes the following result:

THEOREM 1

If $f(z)$ is analytic, the transformation $w = f(z)$ will have a single-valued inverse in the neighborhood of any point where the derivative of the mapping function is different from zero.

Exceptional points where $f'(z) = 0$ are known as **critical points** of the transformation

Now consider a value z and its image $w = f(z)$, where $f(z)$ is analytic, and let

$$\Delta z = |\Delta z|e^{i\theta} \qquad \text{and} \qquad \Delta w = |\Delta w|e^{i\phi}$$

be corresponding increments of these quantities (Fig 17.3) Then

$$f'(z) = \lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z} = \lim_{\Delta z \to 0} \frac{|\Delta w|e^{i\phi}}{|\Delta z|e^{i\theta}} = \lim_{\Delta z \to 0} \left(\frac{|\Delta w|}{|\Delta z|} e^{i(\phi - \theta)}\right)$$

From this it is apparent that

$$\lim_{\Delta z \to 0} \frac{|\Delta w|}{|\Delta z|} = |f'(z)| \qquad \text{and} \qquad \lim_{\Delta z \to 0} (\phi - \theta) = \arg f'(z)$$

or, to an arbitrary degree of approximation,

$$|\Delta w| = |f'(z)|\,|\Delta z| \tag{1}$$

and $\phi = \theta + \arg f'(z)$, or

$$\arg \Delta w = \arg \Delta z + \arg f'(z) \tag{2}$$

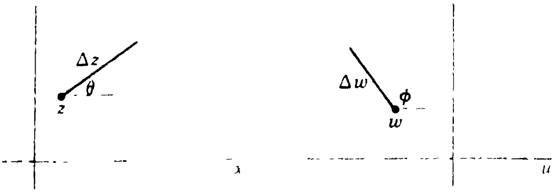


FIGURE 17 3
Plot showing Δz and its image Δw under a mapping $w = f(z)$

---

* See, for instance, R C Buck, "Advanced Calculus," p 215, McGraw-Hill Book Company, New York, 1956.

Now the fact that $f'(z)$ exists [which, of course, it does, since $f(z)$ is assumed to be analytic] means that both $|f'(z)|$ and $\arg f'(z)$ are independent of the manner in which $\Delta z \to 0$. In other words, they depend solely on z and not on the limiting orientation of the increment Δz. Hence, from (1) we draw the following conclusion

THEOREM 2

In the mapping defined by an analytic function $w = f(z)$, the lengths of infinitesimal segments, regardless of their direction, are altered by a factor $|f'(z)|$ which depends only on the point from which the segments are drawn.

Since infinitesimal lengths are magnified by the factor $|f'(z)|$, it follows that infinitesimal areas are magnified by the factor $|f'(z)|^2$, that is, by $J(u,v/x,y)$

Similarly, we conclude from (2) that the difference between the angles of an infinitesimal segment and its image is independent of the direction of the segment and depends only on the point from which the segment is drawn In particular, two infinitesimal segments forming an angle will both be rotated in the same direction by the same amount, hence, the measure of the angle between them will in general be left invariant by the transformation

However, when $f'(z) = 0$, $\arg f'(z)$ is undefined, and we cannot assert that angles are preserved. To investigate this case, suppose that $f'(z)$ has an n-fold zero at $z = z_0$. Then $f'(z)$ must contain the factor $(z - z_0)^n$, and, hence, we can write

$$f'(z) = (n+1)a(z-z_0)^n + (n+2)b(z-z_0)^{n+1} + \cdots$$

where $a, b, \ldots$ are complex coefficients of no concern to us and the factors $n+1, n+2, \ldots$ have been inserted for convenience in integrating $f'(z)$ to obtain $f(z)$.

$$f(z) = f(z_0) + a(z-z_0)^{n+1} + b(z-z_0)^{n+2} + \cdots$$

If in this expression we transpose $f(z_0)$, set

$$z - z_0 = \Delta z \qquad f(z) - f(z_0) = \Delta w$$

and divide by $a(\Delta z)^{n+1}$, we obtain

$$\frac{\Delta w}{a(\Delta z)^{n+1}} = 1 + \frac{b}{a}\Delta z + \cdots$$

As $\Delta z \to 0$, the right member approaches 1 Therefore,

$$\lim_{\Delta z \to 0} (\arg \Delta w) - \lim_{\Delta z \to 0} \arg a(\Delta z)^{n+1} = \arg 1 = 0$$

or, to an arbitrary degree of approximation,

$$\arg \Delta w = \arg a + (n+1) \arg \Delta z$$

Now let Δz_1 and Δz_2 be two infinitesimal segments which make an angle θ with each other, and let Δw_1 and Δw_2 be their images.

From the last expression we have

$$\arg \Delta w_1 = \arg a + (n+1) \arg \Delta z_1$$

$$\arg \Delta w_2 = \arg a + (n+1) \arg \Delta z_2$$

Hence, subtracting,

$$\arg \Delta w_2 - \arg \Delta w_1 = (n+1)(\arg \Delta z_2 - \arg \Delta z_1) = (n+1)\theta$$

Thus we have established the following theorem

THEOREM 3

In the mapping defined by an analytic function $w = f(z)$, angles are in general preserved in magnitude and in sense The only exception to this occurs when the vertex of the angle is an n-fold zero of $f'(z)$, in which case the angle is altered by the factor $n + 1$.

Example 1 of the last section is an excellent illustration of the behavior described by Theorem 3 The mapping function $w = f(z) = z^2$ is everywhere analytic, and, as Fig 17 1 indicates, angles are in general preserved However, the derivative $f'(z) = 2z$ has a simple zero at $z = 0$, and, as Fig 17 2 indicates, angles with vertex at the origin are not preserved, but instead are doubled.

A transformation which preserves the magnitudes of angles is said to be **isogonal.** A transformation which preserves the sense as well as the magnitudes of angles is said to be **conformal.** If $f(z)$ is an analytic function, it follows from Theorem 3 that, in the neighborhood of any point where $f'(z) \neq 0$, the transformation defined by $w = f(z)$ is conformal Conversely, it can be shown* that, if the mapping

$$u = u(x,y) \qquad v = v(x,y)$$

is conformal and if the first partial derivatives of u and v are continuous, then $w = u + iv = f(z)$ is an analytic function. Because of the properties guaranteed by Theorems 2 and 3, it is clear that under a conformal transformation any infinitesimal configuration and its image *conform*, in the sense of being approximately similar. This is not true, however, for large configurations which may bear little or no resemblance to their images

One important reason for studying conformal transformations is that solutions of Laplace's equation remain solutions of Laplace's equation when subjected to a conformal transformation More precisely, we have the following theorem:

THEOREM 4

If $\phi(x,y)$ is a solution of the equation

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0$$

---

* See, for instance, E. G Phillips, "Functions of a Complex Variable," pp 35, 36, Interscience Publishers, Inc , New York, 1945.

then when $\phi(x,y)$ is transformed into a function of u and v by a conformal transformation, it will satisfy the equation

$$\frac{\partial^2\phi}{\partial u^2}+\frac{\partial^2\phi}{\partial v^2}=0$$

everywhere except possibly at the images of the points where the derivative of the mapping function is equal to zero

PROOF Let $w = u(x,y) + iv(x,y)$ define a conformal transformation by means of which $\phi(x,y)$ is transformed into a function of u and v Then

$$\frac{\partial\phi}{\partial x}=\frac{\partial\phi}{\partial u}\frac{\partial u}{\partial x}+\frac{\partial\phi}{\partial v}\frac{\partial v}{\partial x}\qquad\text{and}\qquad\frac{\partial\phi}{\partial y}=\frac{\partial\phi}{\partial u}\frac{\partial u}{\partial y}+\frac{\partial\phi}{\partial v}\frac{\partial v}{\partial y}$$

A second differentiation of each of these yields the results

$$\frac{\partial^2\phi}{\partial x^2}=\frac{\partial\phi}{\partial u}\frac{\partial^2 u}{\partial x^2}+\left(\frac{\partial^2\phi}{\partial u^2}\frac{\partial u}{\partial x}+\frac{\partial^2\phi}{\partial v\,\partial u}\frac{\partial v}{\partial x}\right)\frac{\partial u}{\partial x}+\frac{\partial\phi}{\partial v}\frac{\partial^2 v}{\partial x^2}+\left(\frac{\partial^2\phi}{\partial u\,\partial v}\frac{\partial u}{\partial x}+\frac{\partial^2\phi}{\partial v^2}\frac{\partial v}{\partial x}\right)\frac{\partial v}{\partial x}$$

$$\frac{\partial^2\phi}{\partial y^2}=\frac{\partial\phi}{\partial u}\frac{\partial^2 u}{\partial y^2}+\left(\frac{\partial^2\phi}{\partial u^2}\frac{\partial u}{\partial y}+\frac{\partial^2\phi}{\partial v\,\partial u}\frac{\partial v}{\partial y}\right)\frac{\partial u}{\partial y}+\frac{\partial\phi}{\partial v}\frac{\partial^2 v}{\partial y^2}+\left(\frac{\partial^2\phi}{\partial u\,\partial v}\frac{\partial u}{\partial y}+\frac{\partial^2\phi}{\partial v^2}\frac{\partial v}{\partial y}\right)\frac{\partial v}{\partial y}$$

When these are added, we obtain

$$\begin{aligned}\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}&=\frac{\partial\phi}{\partial u}\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\right)+\frac{\partial^2\phi}{\partial u^2}\left[\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2\right]\\&\quad+2\frac{\partial^2\phi}{\partial u\,\partial v}\left(\frac{\partial u}{\partial x}\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\frac{\partial v}{\partial y}\right)+\frac{\partial\phi}{\partial v}\left(\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}\right)+\frac{\partial^2\phi}{\partial v^2}\left[\left(\frac{\partial v}{\partial x}\right)^2+\left(\frac{\partial v}{\partial y}\right)^2\right]\end{aligned}$$

Since $w = u + iv$ is analytic, by hypothesis, u and v themselves satisfy Laplace's equation Hence, the first and fourth groups of terms on the right vanish identically Moreover, u and v also satisfy the Cauchy-Riemann equations, hence the third group of terms also vanishes identically Using the Cauchy-Riemann equations again, what remains can be written

$$\begin{aligned}\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}&=\frac{\partial^2\phi}{\partial u^2}\left[\left(\frac{\partial u}{\partial x}\right)^2+\left(-\frac{\partial v}{\partial x}\right)^2\right]+\frac{\partial^2\phi}{\partial v^2}\left[\left(\frac{\partial v}{\partial x}\right)^2+\left(\frac{\partial u}{\partial x}\right)^2\right]\\&=\left[\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial v}{\partial x}\right)^2\right]\left(\frac{\partial^2\phi}{\partial u^2}+\frac{\partial^2\phi}{\partial v^2}\right)\\&=|f'(z)|^2\left(\frac{\partial^2\phi}{\partial u^2}+\frac{\partial^2\phi}{\partial v^2}\right)\end{aligned}$$

Thus, at any point where the transformation is conformal, that is, where $f'(z) \neq 0$,

$$\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}=0\qquad\text{implies}\qquad\frac{\partial^2\phi}{\partial u^2}+\frac{\partial^2\phi}{\partial v^2}=0\qquad\text{as asserted}$$

Suppose now that it is required to solve Laplace's equation, subject to certain boundary conditions, within a region R. Unless R is of a very simple shape, a direct attack upon the problem will usually be exceedingly difficult. However, it may be possible to find a conformal transformation which will convert R into some simpler region R', such as a circle or a half plane, in which Laplace's equation can be solved, subject, of course, to the transformed boundary conditions. If this is the case, the resulting solution, when carried back to R by the inverse transformation, will be the required solution of the original problem.

EXERCISES

1 **a** What is the length of the curve into which the upper half of the circle $|z| = a$ is transformed by the function $w = 1/z$? (**b**) What is the length of the arc into which this function transforms the segment of $y = 1 - x$ which lies in the first quadrant?

2 What is the area of the region into which the square with vertices $z = 0, 1, 1 + i, i$ is transformed (**a**) by the function $w = z^2$? (**b**) by $w = z^3$?

3 **a** What are the critical points of the transformation $w = 3z - z^3$? (**b**) What is the locus of points at which the magnification is 1? What is the locus of points at which infinitesimal segments are rotated (**c**) through 45°? (**d**) through 90°?

4 Are there any points at which infinitesimal segments are left unchanged in magnitude and direction by the transformation $w = z^2 + z^3$?

5 If $u = 2x^2 + y^2$ and $v = y^2/x$, show that the curves $u =$ constant and $v =$ constant cut orthogonally at all intersections, but that the transformation defined by $f(z) = u + iv$ is not conformal Give a specific illustration of the latter fact

17.3

The bilinear transformation

The simplest class of conformal transformations, yet one of the most important, is the class of **bilinear** or **linear fractional** or **Möbius transformations,*** defined by the family of functions

$$w = \frac{az + b}{cz + d} \qquad ad - bc \neq 0 \tag{1}$$

The restriction $ad - bc \neq 0$ is necessary because, if $ad = bc$, then $a/c = b/d$ and the numerator and denominator of w are proportional As a consequence, w is a constant independent of z, and thus the entire z-plane is mapped into the same point in the w-plane!

It is convenient to investigate the general bilinear transformation by considering first the three special cases

a $w = z + \lambda$
b $w = \mu z$
c $w = 1/z$

In case **a**, w is found by adding a constant vector λ to each z. Hence the transformation is just a translation in the direction defined by $\arg \lambda$ through a distance equal to $|\lambda|$ In particular, we note for later use that this rigid motion necessarily transforms circles into circles.

In case **b**, w is found by rotating each z through a fixed angle equal to $\arg \mu$ and then multiplying its length by the factor $|\mu|$.

---

* Named for the German geometer A F. Möbius (1790–1868).

In this case, too, circles are transformed into circles. To prove this, let us first write the equation of the general circle

$$a(x^2 + y^2) + bx + cy + d = 0 \qquad a, b, c, d \text{ real} \qquad b^2 + c^2 \geqq 4ad$$

in terms of z and $\bar{z}$ by means of the relations

$$x = \frac{z + \bar{z}}{2} \qquad y = \frac{z - \bar{z}}{2i} \qquad x^2 + y^2 = z\bar{z}$$

The result is $\quad az\bar{z} + \dfrac{b - ic}{2} z + \dfrac{b + ic}{2} \bar{z} + d = 0$

or, renaming the coefficients,

(2) $$(A + \bar{A})z\bar{z} + Bz + \bar{B}\bar{z} + (D + \bar{D}) = 0$$

where now A, B, and D can be arbitrary complex numbers, subject to the condition $B\bar{B} \geqq (A + \bar{A})(D + \bar{D})$, derived from the condition $b^2 + c^2 \geqq 4ad$, which ensures that the radius of the circle is real. If the substitution

$$z = \frac{w}{\mu}$$

is made in (2), we obtain the transformed equation

$$(A + \bar{A})\frac{w}{\mu}\frac{\bar{w}}{\bar{\mu}} + B\frac{w}{\mu} + \bar{B}\frac{\bar{w}}{\bar{\mu}} + (D + \bar{D}) = 0$$

or

(3) $$(A + \bar{A})w\bar{w} + (B\bar{\mu})w + (\bar{B}\mu)\bar{w} + (D + \bar{D})\mu\bar{\mu} = 0$$

Since the coefficients of the first and last terms in (3) are real and since the coefficients of w and $\bar{w}$ are conjugates, this equation has the same structure as (2) and, hence, will also represent a circle provided its coefficients satisfy the condition necessary for the radius to be real For (3), this condition is

$$(B\bar{\mu})(\bar{B}\mu) \geqq (A + \bar{A})(D + \bar{D})\mu\bar{\mu}$$

or, dividing through by $\mu\bar{\mu}$, which is necessarily positive,

$$B\bar{B} \geqq (A + \bar{A})(D + \bar{D})$$

which is true by hypothesis. If $a = 0$, so that $A + \bar{A} = 0$, both the given circle and its image reduce to straight lines

In case c we can write

(4) $$w = \frac{1}{z} = \frac{\bar{z}}{z\bar{z}}$$

which shows that w is of length $1/|z|$ and has the direction of $\bar{z}$.

To describe the geometrical process by which a point with these characteristics can be obtained from a given point z, we must first define the process of **inversion**. Let C be a circle with center O and radius r, and let P be any point in the plane of C Then the **inverse** of P with respect to C is the point P' on the ray OP for which

(5) $$OP \cdot OP' = r^2$$

From the symmetry of this relation it is clear that P is also the inverse of P'. Geometrically, a point and its inverse are related as follows. From any point P outside a circle C with center O let the two tangents to C be drawn, and let the points of contact of these tangents be joined (Fig. 17.4). The intersection of this chord with the line OP is the inverse P' of P. Conversely, let P' be any point in the interior of C. At P' erect a perpendicular to OP', and at the point where this meets C let the tangent to C be drawn. The intersection of this tangent and the line OP' is the inverse P of P'. The consistency of these constructions with the definitive property (5) is evident, since in Fig. 17.4

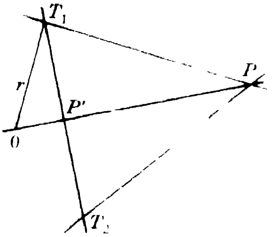


FIGURE 17.4
Plot showing the geometrical relation between a point and its inverse

$$\Delta OP'T_1 \sim \Delta OT_1P$$

and thus $$\frac{OP'}{OT_1} = \frac{OT_1}{OP}$$

or $$OP \cdot OP' = (OT_1)^2 = r^2$$

It is evident now that the construction of w from z in case c requires that the inverse of z in the unit circle be found and then reflected in the real axis; for the first of these steps gives a complex number whose length is $1/|z|$, and the second achieves the direction of $\bar{z}$, as required by (4).

To show that circles are also transformed into circles in case c, let the substitution $z = 1/w$ be made in the self-conjugate form of the equation of a circle (2). This gives

$$(A + \bar{A})\frac{1}{w}\frac{1}{\bar{w}} + \frac{B}{w} + \frac{\bar{B}}{\bar{w}} + (D + \bar{D}) = 0$$

or $$(D + \bar{D})w\bar{w} + \bar{B}w + B\bar{w} + (A + \bar{A}) = 0$$

which is also the equation of a circle with real radius. If $A + \bar{A} = 0$, the original circle reduces to a straight line whose image is a circle passing through the origin, since its equation contains no constant term. Conversely, any circle passing through the origin is transformed into a straight line.

The three special transformations we have just considered can be used to synthesize the general bilinear transformation. To see this, suppose first that $c \neq 0$. Then the general transformation is equivalent to the following chain of special transformations:

$$
\begin{aligned}
w_1 &= z + \frac{d}{c} \\
w_2 &= cw_1 = cz + d \\
w_3 &= \frac{1}{w_2} = \frac{1}{cz + d} \\
w_4 &= \frac{bc - ad}{c} w_3 = \frac{bc - ad}{c(cz + d)} \\
w &= w_4 + \frac{a}{c} = \frac{bc - ad}{c(cz + d)} + \frac{a}{c} = \frac{az + b}{cz + d}
\end{aligned}
$$

On the other hand, if $c = 0$, it is clear from the restriction $ad - bc \neq 0$ that neither a nor d can be zero. Hence, we can write

$$
\begin{aligned}
w_1 &= z + \frac{b}{a} \\
w &= \frac{a}{d} w_1 = \frac{a}{d}\left(z + \frac{b}{a}\right) = \frac{az + b}{d}
\end{aligned}
$$

Thus we have shown that in all cases the general bilinear transformation can be compounded from a succession of simple transformations of types **a**, **b**, and **c**. Since each of these is known to transform circles into circles, including straight lines as special cases, we have thus established the following theorem.

THEOREM 1

Under the general bilinear transformation circles are transformed into circles.

The general bilinear transformation

$$
w = \frac{az + b}{cz + d}
$$

depends on three essential constants, namely, the ratios of any three of the constants a, b, c, d to the fourth. Hence it is evident that three conditions are necessary to determine a bilinear transformation. In particular, the requirement that three distinct values of z, say z_1, z_2, z_3, have specified distinct images w_1, w_2, w_3 leads to a unique transformation.

Although the transformation which sends three given points into three specified image points can be found by imposing these conditions on the general equation and solving for the constants, it is generally simpler to make use of the fact that if w_1, w_2, w_3, w_4 are, respectively, the images of z_1, z_2, z_3, z_4, then

$$
\frac{(w_1 - w_2)(w_3 - w_4)}{(w_1 - w_4)(w_3 - w_2)} = \frac{(z_1 - z_2)(z_3 - z_4)}{(z_1 - z_4)(z_3 - z_2)}
$$

To establish this relation, we observe that

$$
w_i - w_j = \frac{az_i + b}{cz_i + d} - \frac{az_j + b}{cz_j + d} = \frac{(ad - bc)(z_i - z_j)}{(cz_i + d)(cz_j + d)}
$$

Hence

$$\frac{(w_1 - w_2)(w_3 - w_4)}{(w_1 - w_4)(w_3 - w_2)} = \frac{\dfrac{(ad - bc)(z_1 - z_2)}{(cz_1 + d)(cz_2 + d)}\,\dfrac{(ad - bc)(z_3 - z_4)}{(cz_3 + d)(cz_4 + d)}}{\dfrac{(ad - bc)(z_1 - z_4)}{(cz_1 + d)(cz_4 + d)}\,\dfrac{(ad - bc)(z_3 - z_2)}{(cz_3 + d)(cz_2 + d)}}$$

$$= \frac{(z_1 - z_2)(z_3 - z_4)}{(z_1 - z_4)(z_3 - z_2)}$$

The last fraction is called the **cross ratio or anharmonic ratio** of the four numbers z_1, z_2, z_3, z_4; hence the result we have just established can be formulated as the following theorem.

THEOREM 2

The cross ratio of four points is invariant under a bilinear transformation

Suppose now that it is required to find the transformation which sends z_1, z_2, z_3 into w_1, w_2, w_3, respectively. If w is the image of a general point z under this transformation, then, according to Theorem 2, the cross ratio of w_1, w_2, w_3, and w must equal the cross ratio of z_1, z_2, z_3, and z That is,

$$\frac{(w_1 - w_2)(w_3 - w)}{(w_1 - w)(w_3 - w_2)} = \frac{(z_1 - z_2)(z_3 - z)}{(z_1 - z)(z_3 - z_2)}$$

This equation is clearly bilinear in w and z and is satisfied by the three pairs of values (z_1,w_1), (z_2,w_2), (z_3,w_3) Moreover, everything in it is known except the variables w and z themselves; hence it is necessary only to solve for w in terms of z to obtain the required transformation in standard form.

EXAMPLE 1

What is the bilinear transformation which sends the points $z = -1, 0, 1$ into the points $w = 0, i, 3i$, respectively?

Setting up the appropriate cross ratios, we have

$$\frac{(0 - i)(3i - w)}{(0 - w)(3i - i)} = \frac{(-1 - 0)(1 - z)}{(-1 - z)(1 - 0)}$$

or

$$\frac{3i - w}{2w} = \frac{1 - z}{1 + z}$$

Solving for w, we obtain without difficulty

$$w = -3i\frac{z + 1}{z - 3}$$

EXAMPLE 2

What is the most general bilinear transformation which maps the upper half of the z-plane onto the interior of the unit circle in the w-plane (Fig. 17 5)?

Let the required transformation be

$$w = \frac{az + b}{cz + d}$$

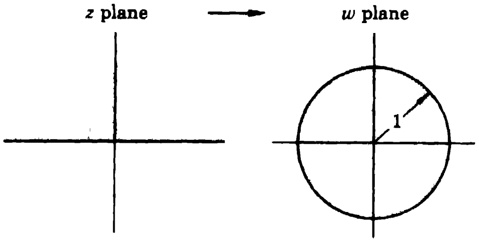


FIGURE 17.5 The upper half of the z-plane to be mapped onto the interior of the unit circle in the w-plane

Since the boundaries of corresponding regions must correspond under any transformation, the unit circle in the w-plane must be the image of the real axis in the z-plane. Therefore, for all real values of z, we must have

$$|w| = \frac{|az + b|}{|cz + d|} = \frac{|a|}{|c|}\,\frac{|z + (b/a)|}{|z + (d/c)|} = 1$$

In particular, from the limiting case $|z| \to \infty$, we find

$$\frac{|a|}{|c|} = 1$$

and thus for real values of z,

$$\left|z + \frac{b}{a}\right| = \left|z + \frac{d}{c}\right| \qquad \text{or} \qquad \left|z - \left(-\frac{b}{a}\right)\right| = \left|z - \left(-\frac{d}{c}\right)\right|$$

The last equation expresses the fact that the complex numbers $-b/a$ and $-d/c$ are equally far from all points on the real axis, which is possible if and only if the real axis is the perpendicular bisector of the segment joining the points $-b/a$ and $-d/c$. Therefore, $-b/a$ and $-d/c$ must be conjugates, say λ and $\bar\lambda$. Thus we can write

$$\begin{aligned} w &= \frac{az + b}{cz + d} \\ &= \frac{a}{c}\,\frac{z + (b/a)}{z + (d/c)} \\ &= \frac{a}{c}\,\frac{z - \lambda}{z - \bar\lambda} \\ &= e^{i\theta}\,\frac{z - \lambda}{z - \bar\lambda} \qquad (6) \end{aligned}$$

where the last step follows because, as we found earlier, a/c is a complex number of absolute value 1.

So far we have enforced only the condition that the boundaries of the two regions correspond. It is now necessary to make sure that the regions themselves correspond as required and that the upper half of the z-plane has not been mapped into the *outside* of the circle $|w| = 1$. This is most easily verified by checking some convenient point, say $z = \lambda$. This maps into $w = 0$, which is certainly inside the circle $|w| = 1$. Thus, if λ is restricted to lie in the *upper* half of the z-plane, the solution is complete.

As a special case of some interest, let $e^{i\theta} = -1$, and let λ be a pure imaginary, say i. Then

$$(7) \qquad w = -\frac{z - i}{z + i}$$

Now

$$\mathscr{I}(w) = \frac{w - \bar w}{2i} = -\frac{1}{2i}\left(\frac{z - i}{z + i} - \frac{\bar z + i}{\bar z - i}\right)$$

or, reducing to a common denominator and simplifying,

$$\mathscr{I}(w) = \frac{z + \bar{z}}{(z + i)(\bar{z} - i)}$$

The denominator of the last fraction is the product of $z + i$ and its conjugate $\bar{z} - i$ and, hence, is a positive quantity. Thus, the imaginary part of w will be positive if and only if $z + \bar{z}$ is positive. Since $z + \bar{z}$ is equal to twice the real part of z, this shows that the transformation (7) not only maps the upper half of the z-plane onto the unit circle $|w| \leq 1$ but does it in such a way that the first quadrant of the z-plane [where $\mathscr{R}(z) > 0$] corresponds to the upper half of the circle [where $\mathscr{I}(w) > 0$] and the second quadrant of the z-plane corresponds to the lower half of the circle. In the opposite direction, the inverse transformation

$$z = -i\frac{w - 1}{w + 1} \tag{8}$$

maps the interior of the circle $|w| = 1$ onto the upper half of the z-plane in such a way that the upper half of the circle maps onto the first quadrant of the z-plane.

EXAMPLE 3

Find a transformation which will map an infinite sector of angle $\pi/4$ onto the interior of the unit circle.

Since the boundary of the sector consists of portions of two straight lines, yet its image is to be a single circle, it is apparent that the mapping cannot be accomplished by a bilinear transformation alone. However, a simple combination of a power function and a linear fractional function will define a suitable transformation. Specifically, the transformation

$$t = z^4$$

will open out the sector in the z-plane into the upper half of the auxiliary t-plane (Fig. 17.6).

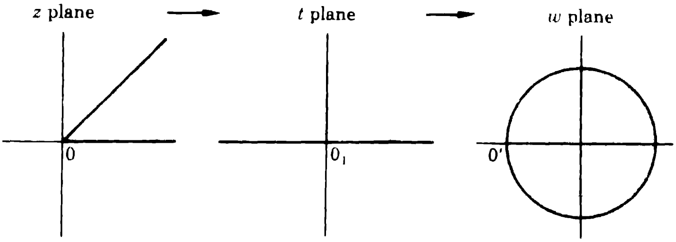


FIGURE 17.6 The two transformations needed to map an infinite sector onto the interior of the unit circle

Following this, the upper half of the t-plane can be mapped onto the unit circle in the w-plane by any transformation of the family (6), which we obtained in the last example, say

$$w = \frac{t - i}{t + i}$$

Combining these two, we have for the required transformation

$$w = \frac{z^4 - i}{z^4 + i}$$

EXAMPLE 4

Find a transformation which will map a 60° sector of the unit circle in the z-plane onto the upper half of the w-plane.

At first glance it would seem that this problem can be solved simply by opening the given sector into a full circle by the transformation

$$t = z^6$$

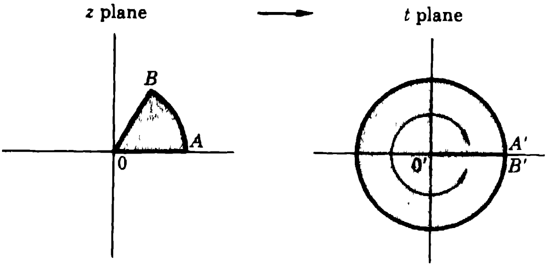


FIGURE 17.7
A circular sector "opened out" into a circular region cut along a radius

and then mapping the circle from the t-plane onto the upper half of the w-plane by means of the inverse of one of the transformations of the family (6) which we obtained in Example 2, for instance, the transformation (8). This method fails, however, because the circular region obtained in the t-plane in this case is *not* of the type considered in Example 2. The latter consisted of a simple circular boundary plus its interior, whereas the former consists of the interior of a circle "cut" along a radius, since the radius $O'A' \equiv O'B'$ is actually the image of the two boundary radii OA and OB (Fig. 17.7).

To avoid this difficulty, let us first map the sector onto a semicircle by the transformation

$$t_1 = z^3$$

Then let us map the semicircle from the t_1-plane onto the first quadrant of the t_2-plane by means of the transformation (8)

$$t_2 = -i\frac{t_1 - 1}{t_1 + 1}$$

Finally (Fig. 17.8) let us open out the first quadrant of the t_2-plane into the upper half of the w-plane by the transformation

$$w = t_2^2$$

Combining these three transformations, we find

$$w = -\left(\frac{z^3 - 1}{z^3 + 1}\right)^2$$

as the required solution

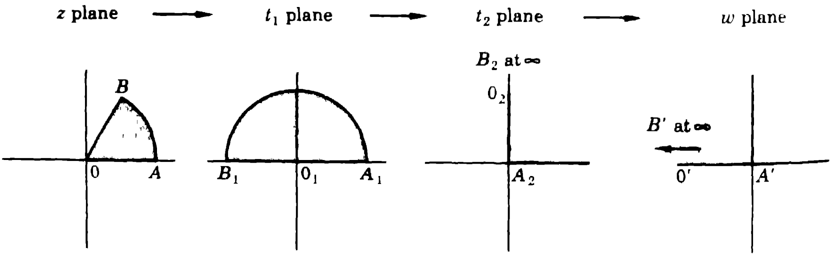


FIGURE 17.8
The sequence of transformations necessary to map a circular sector onto a half plane

EXAMPLE 5

A thin sheet of metal coincides with the first quadrant of the z-plane. The upper and lower faces of the sheet are perfectly insulated against the flow of heat. Find the steady-state temperature at any point of the sheet if the boundary temperatures are those shown in Fig. 17.9a

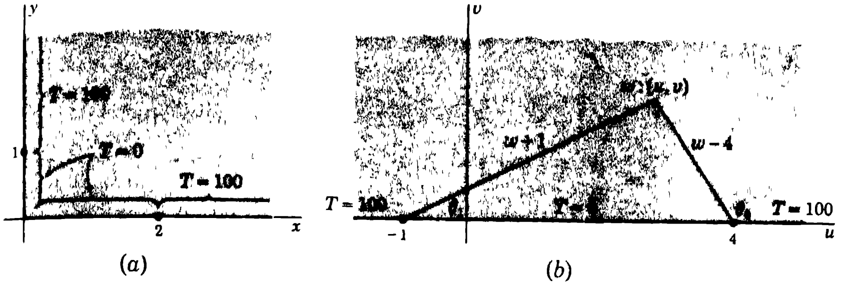


FIGURE 17.9

An infinite 90° sector mapped, with its boundary conditions, onto a half plane

Under the assumptions of the problem, the flow of heat in the sheet is two-dimensional, and we must accordingly solve Laplace's equation, i.e., the two-dimensional steady-state heat equation derived in Sec. 8.2,

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0$$

subject to the given conditions along the boundaries of the first quadrant. To do this, it is convenient to map the first quadrant of the z-plane onto the upper half of the w-plane by the transformation

$$w = z^2 = (x^2 - y^2) + 2ixy$$

This reduces the problem to that of finding a solution of Laplace's equation in the upper half plane which assumes along the real axis the boundary conditions shown in Fig. 17.9b.

Now we have long since discovered (Property 1, Sec. 14.6) that either the real or the imaginary part of any analytic function satisfies Laplace's equation. In particular, since the function

$$(9)\qquad iT_0 + \frac{1}{\pi}[(T_1 - T_0)\ln(z - x_0) + (T_2 - T_1)\ln(z - x_1) + \\ + (T_{n+1} - T_n)\ln(z - x_n)]$$

is analytic except at the real points $x_0, x_1, \ldots, x_n$, its imaginary part, namely,

$$(10)\qquad T = T_0 + \frac{1}{\pi}[(T_1 - T_0)\arg(z - x_0) + (T_2 - T_1)\arg(z - x_1) + \\ + (T_{n+1} - T_n)\arg(z - x_n)]$$

will be a solution of Laplace's equation. Moreover, along the real axis, this solution takes on the boundary values shown in Fig. 17.10. To see this, we observe from Fig. 17.10 that the complex

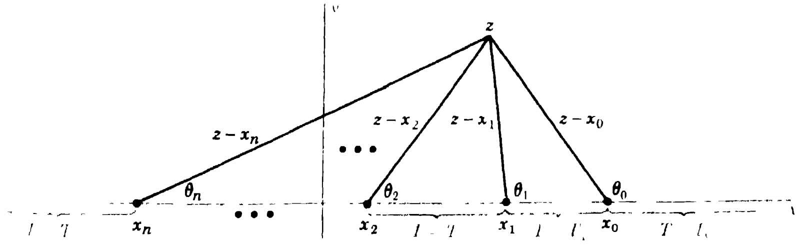


FIGURE 17.10

Plot showing the behavior of arg $(z - x_i)$ as z varies along the real axis number $z - x_i$ is represented by the vector joining the fixed point x_i to the variable point z, and thus arg $(z - x_i)$ is simply the inclination angle of this vector Hence the function (10) can be rewritten

$$(11) \qquad T = T_0 + \frac{1}{\pi}[(T_1 - T_0)\theta_0 + (T_2 - T_1)\theta_1 + \qquad + (T_{n+1} - T_n)\theta_n]$$

Again referring to Fig 17 10, it is clear that, for all values of z on the real axis to the right of x_0, each of the θ's is zero Hence from (11) we see that T reduces to the constant value T_0 along this portion of the real axis Furthermore, when z lies between x_1 and x_0, θ_0 is equal to π, but all the other θ's are still zero Hence, along this segment the temperature (10), or (11), reduces to

$$T = T_0 + \frac{1}{\pi}[(T_1 - T_0)\pi] = T_1$$

Similarly, for values of z between x_2 and x_1, the angles θ_0 and θ_1 are each equal to π, but all other θ's are zero Hence, along this segment, we have

$$T = T_0 + \frac{1}{\pi}[(T_1 - T_0)\pi + (T_2 - T_1)\pi] = T_2$$

Continuing in this fashion, we can verify that T, as defined by (10) or (11), not only is a solution of Laplace's equation, being the imaginary part of the analytic function (9), but also assumes along the real axis the temperature distribution shown in Fig 17 10

Specializing these observations to our problem, it appears that the solution we require is

$$\begin{aligned} T &= 100 + \frac{1}{\pi}[(0 - 100)\theta_0 + (100 - 0)\theta_1] \\ &= 100 + \frac{100}{\pi}(\theta_1 - \theta_0) \\ &= \frac{100}{\pi}[\pi + (\theta_1 - \theta_0)] \end{aligned}$$

Now, multiplying by $\pi/100$ and then taking the tangent of both sides of the last equation, we have

$$\begin{aligned} \tan\frac{\pi T}{100} &= \tan[\pi + (\theta_1 - \theta_0)] = \tan(\theta_1 - \theta_0) \\ &= \frac{\tan\theta_1 - \tan\theta_0}{1 + \tan\theta_0\tan\theta_1} \end{aligned}$$

Substituting for $\tan\theta_0$ and $\tan\theta_1$ their values as read from Fig 17 9b, we obtain from the last expression

$$\begin{aligned} \tan\frac{\pi T}{100} &= \frac{v/(u+1) - v/(u-4)}{1 + v^2/(u+1)(u-4)} \\ (12) \qquad &= \frac{-5v}{u^2 + v^2 - 3u - 4} \end{aligned}$$

which is the solution of the transformed problem in the w-plane Returning to the z-plane by means of the transformation equations

$$u = x^2 - y^2 \qquad \text{and} \qquad v = 2xy$$

we thus find, from (12), that

$$T = \frac{100}{\pi}\tan^{-1}\frac{-10xy}{(x^2 + y^2)^2 - 3x^2 + 3y^2 - 4}$$

is the solution to the original problem

EXERCISES

1 What is the cross ratio of the four fourth roots of -1?

2 What is the cross ratio of the four complex sixth roots of 1?

3 Show that in general there are two points which are left invariant by a bilinear transformation. Are there any bilinear transformations which leave only one point invariant? no points invariant?

4 Find the invariant points of the transformation $w = -(2z + 4i)/(iz + 1)$, and prove that these two points, together with any point z and its image w, form a set of four points having a constant cross ratio.

5 What is the bilinear transformation which sends the points $z = 0, -1, \infty$ into the points $w = -1, -2 - i, i$, respectively? What is the image of the circle $|z| = 1$ under this transformation?

6 What is the bilinear transformation which sends the points $z = 0, -i, 2i$ into the points $w = 5i, \infty, -i/3$, respectively? What are the invariant points of this transformation?

7 What is the most general bilinear transformation which maps the upper half of the z-plane onto the lower half of the w-plane?

8 Prove that $w = z/(1 - z)$ maps the upper half of the z-plane into the upper half of the w-plane. What is the image of the circle $|z| = 1$ under this transformation?

9 Find a transformation which will map an infinite sector of angle $\pi/3$ onto the interior of the unit circle.

10 Show that along the circle $|cz + d| = \sqrt{|ad - bc|}$ the transformation $w = (az + b)/(cz + d)$ does not alter the lengths of infinitesimal segments. What happens to segments inside this circle? outside this circle? What is the locus of points where infinitesimal segments are not rotated by the transformation?

11 Find a transformation which will map a 45° sector of the unit circle in the z-plane onto the upper half of the w-plane.

12 Find a transformation which will map the upper half of the unit circle onto the entire unit circle.

13 Show that, if $|c| = |d|$, then the transformation $w = (az + b)/(cz + d)$ maps the unit circle in the z-plane into a straight line in the w-plane.

14 Verify that the transformation $w = re^{i\alpha}(z - z_1)/(z - z_2)$ maps the region in the z-plane bounded by two circular arcs intersecting at an angle α at z_1 and z_2 into the interior of an angle α in standard position in the w-plane. [Hint: Recall that an equation of the form (2) represents a circle in the z-plane.]

15 Prove that four points z_1, z_2, z_3, z_4 lie on a circle if and only if their cross ratio is real.

16 Find the steady-state temperature distribution in a sheet of metal coinciding with the first quadrant of the z-plane if $T = 100°$ along the positive x-axis and $T = 0°$ along the positive y-axis.

17 Find the steady-state temperature distribution in a sheet of metal coinciding with the interior of a 60° angle in standard position in the z-plane if $T = 0°$ along the horizontal side of the angle and $T = 100°$ along the other side.

18 Find the steady-state temperature distribution in a sheet of metal coinciding with the first quadrant of the z-plane if $T = 100°$ along the positive y-axis, if $T = 50°$ between 0 and 3 on the x-axis, and if $T = 0$ to the right of 3 on the x-axis.

19 Find the steady-state temperature distribution in the unit circle in the z-plane if the upper half of the boundary of the circle is kept at the temperature $T = 100°$ and the lower half of the boundary is kept at the temperature $T = 0°$.

20 Show that $w = z + 1/z$ maps the portion of the upper half of the z-plane exterior to the circle $|z| = 1$ onto the entire upper half of the w-plane. Use this result to find the steady-state temperature distribution in the upper half of the z-plane exterior to the unit circle if $T = 100°$ along the linear portion of the boundary and $T = 0°$ along the circular portion of the boundary.

17.4

The Schwarz-Christoffel transformation

In general, the conformal transformation of one given region into another is exceedingly difficult The *existence* of such a transformation is assured by the following theorem, due to Riemann.

THEOREM 1

Any two bounded simply connected regions can be mapped conformally onto each other.

However, the determination of the specific function which accomplishes a required mapping is usually out of the question. In fact, in addition to the simple regions which we found could be mapped by means of the elementary functions, the only class of regions for which conformal transformations of practical interest exist are those bounded by polygons having a finite number of vertices (one or more of which may lie at infinity). These can always be mapped onto a half plane and, hence, onto any region into which a half plane can be transformed, by means of a transformation which we shall now discuss

To see how this can be done, we first recall the mapping properties of the power function

$$w = z^m$$

Since this transformation has the property (Theorem 3, Sec. 17 2) that it alters by the factor m any angle with vertex at the origin, it follows that the transformation

(1) $$w - w_1 = (z - x_1)^{\alpha_1/\pi}$$

will take a segment of the x-axis containing x_1 in its interior, i e , a straight angle with vertex at x_1, and "fold" it into an angle of

$$\frac{\alpha_1}{\pi}\pi = \alpha_1$$

with vertex at w_1. Clearly, if this could be done simultaneously for a number of points $x_1, x_2, \ldots, x_n$ on the x-axis, the x-axis would be mapped into a polygon whose angles were, respectively, $\alpha_1, \alpha_2, \ldots, \alpha_n$ and conversely, and the biggest step in the solution of our problem would be taken This is actually possible, and the transformation which accomplishes it, suggested by the form of the derivative of (1), is defined by

(2) $$\frac{dw}{dz} = K(z - x_1)^{(\alpha_1/\pi)-1}(z - x_2)^{(\alpha_2/\pi)-1} \cdots (z - x_n)^{(\alpha_n/\pi)-1}$$

To verify this, we begin with a point z on the x-axis to the left of the first of the given points $x_1, x_2, \ldots, x_n$ and investigate the locus of its image as it moves to the right along the x-axis (Fig.

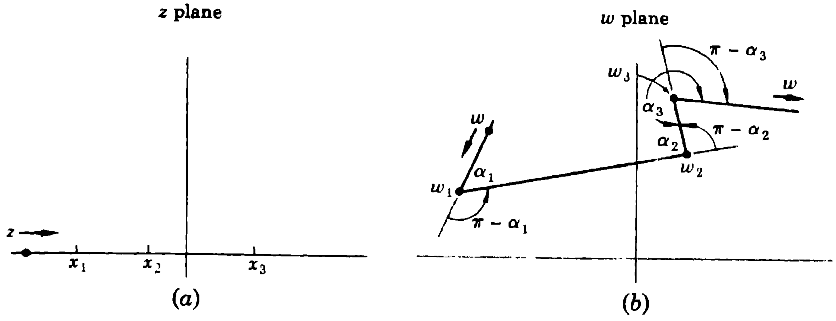


FIGURE 17.11

The mapping of the real axis in the z-plane into a polygon with prescribed angles in the w-plane.

17.11) From (2) we obtain at once the relation

$$\begin{aligned}\arg dw = \arg K &+ \left(\frac{\alpha_1}{\pi} - 1\right)\arg(z - x_1) \\ &+ \left(\frac{\alpha_2}{\pi} - 1\right)\arg(z - x_2) + \\ &+ \left(\frac{\alpha_n}{\pi} - 1\right)\arg(z - x_n) + \arg dz\end{aligned} \tag{3}$$

and in this it is apparent that until z reaches x_1, every term on the right remains constant, since $z - x_1, z - x_2, \ldots, z - x_n$ are all negative real numbers and, hence, have π for their respective arguments, and since dz is positive and, therefore, has 0 as its argument Thus the image point w traces a straight line, since the argument of the increment dw remains constant However, as z passes through x_1, the difference $z - x_1$ changes abruptly from negative to positive, and thus $\arg(z - x_1)$ decreases abruptly from π to 0 Hence, $\arg dw$ changes by the amount

$$\left(\frac{\alpha_1}{\pi} - 1\right)(-\pi) = \pi - \alpha_1$$

But, from Fig 17.11b, it is evident that this is the precise amount through which it is necessary to turn if w is to begin to move in the direction of the next side of the polygon. As z moves from x_1 to x_2, the same situation exists The argument of dw remains constant, and thus w moves in a straight line until z reaches x_2. Here $z - x_2$ changes abruptly from negative to positive, $\arg(z - x_2)$ jumps from π to 0, and, as a consequence, $\arg dw$ increases by the amount $\pi - \alpha_2$, which is the exact amount of rotation required to give the direction of the next side of the polygon.

Thus as z traverses the x-axis, it is clear that w moves along the boundary of a polygon whose interior angles are precisely the given angles $\alpha_1, \alpha_2, \ldots, \alpha_n$. Moreover, it is evident that the region which is mapped onto the half plane is the region which contains these angles. The required transformation will be obtained if we can ensure that the lengths of the sides of the polygon, as well as its angles, have the correct values.

Now the mapping function w, obtained by integrating (2), is

$$w = K\int[(z - x_1)^{(\alpha_1/\pi)-1}(z - x_2)^{(\alpha_2/\pi)-1} \cdots (z - x_n)^{(\alpha_n/\pi)-1}]\,dz + C \tag{4}$$

and this can be thought of as the result of the two transformations

$$t = \int[(z - x_1)^{(\alpha_1/\pi)-1}(z - x_2)^{(\alpha_2/\pi)-1} \cdots (z - x_n)^{(\alpha_n/\pi)-1}]\,dz \tag{5}$$

$$w = Kt + C \tag{6}$$

The first of these transforms the x-axis into some polygon which the second then translates, rotates, and either stretches or shrinks, as the case may be. If, then, the polygon determined by (5) is similar to the given polygon, the constants in (6) can always be determined so as to make the two polygons coincide.

Now for two polygons to be similar, not only must corresponding angles be equal but corresponding sides must be proportional. For triangles this is automatically the case. For quadrilaterals one further condition is required, namely, that two pairs of corresponding sides have the same ratio. For pentagons two such conditions are required, and, in general, for a polygon of n sides, $n - 3$ conditions, over and above the equality of corresponding angles, are necessary for similarity. Hence, in mapping a polygon of n sides onto a half plane, three of the image points $x_1, x_2, \ldots, x_n$ can be assigned arbitrarily, following which the remaining $n - 3$ are determined by the conditions of similarity. In many important problems, a vertex of the polygon, usually an infinite one, will correspond to $z = \infty$. In this case dw/dz contains one less term than usual and, hence, one less parameter. Therefore, only two of the $n - 1$ finite image points $x_1, x_2, \ldots, x_{n-1}$ can be specified arbitrarily. In either case the resulting transformation is known as the **Schwarz-Christoffel transformation.*** Obviously, since w is analytic everywhere, except possibly at the points $x_1, x_2, \ldots, x_n$, the transformation is conformal. In practice, the usefulness of the Schwarz-Christoffel transformation is often limited by the complexity of the integral which defines the mapping function w.

EXAMPLE 1

Find the transformation which maps the semi-infinite strip shown in Fig. 17.12a onto the half plane, as indicated.

The required transformation is defined by

$$\frac{dw}{dz} = K(z + 1)^{\frac{\pi/2}{\pi}-1}(z - 1)^{\frac{\pi/2}{\pi}-1} = K(z + 1)^{-\frac{1}{2}}(z - 1)^{-\frac{1}{2}}$$

Hence,
$$w = K\int\frac{dz}{\sqrt{z^2 - 1}} = K\cosh^{-1} z + C$$

---

* Named for the German mathematicians H. A. Schwarz (1843–1921) and E. B. Christoffel (1829–1900), who discovered it independently about 1865.

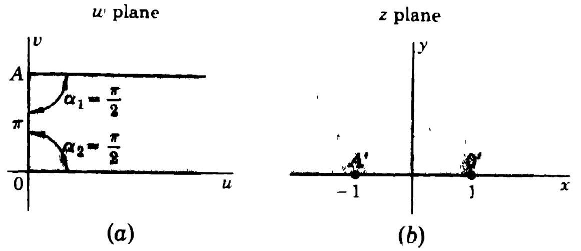


FIGURE 17.12
A semi-infinite strip to be mapped onto a half plane

Since $w = 0$ is to correspond to $z = 1$, we have

$$0 = K \cosh^{-1} 1 + C \qquad \text{or} \qquad C = 0$$

Also, $w = i\pi$ is to correspond to $z = -1$, and thus

$$i\pi = K \cosh^{-1} (-1) = K(i\pi) \qquad \text{or} \qquad K = 1$$

The required transformation is, therefore, $w = \cosh^{-1} z$, or

$$z = \cosh w$$

Broken down into real and imaginary parts, this becomes

$$x + iy = \cosh u \cos v + i \sinh u \sin v$$

or
$$x = \cosh u \cos v$$
$$y = \sinh u \sin v$$

Eliminating u and v in turn, we have also

$$\frac{x^2}{\cosh^2 u} + \frac{y^2}{\sinh^2 u} = 1$$

$$\frac{x^2}{\cos^2 v} - \frac{y^2}{\sin^2 v} = 1$$

which, if necessary, can be solved for u and v in terms of x and y

EXAMPLE 2

Find the transformation which maps the infinite region shown in Fig 17 13*a* onto the upper half plane as indicated

With images assigned as shown and with the angle at the finite vertex A identified as $\alpha_1 = 2\pi$ and the angle at the infinite vertex B identified as $\alpha_2 = 0$, we have

$$\frac{dw}{dz} = K(z+1)^{(2\pi/\pi)-1} z^{(0/\pi)-1} = K\left(1 + \frac{1}{z}\right)$$

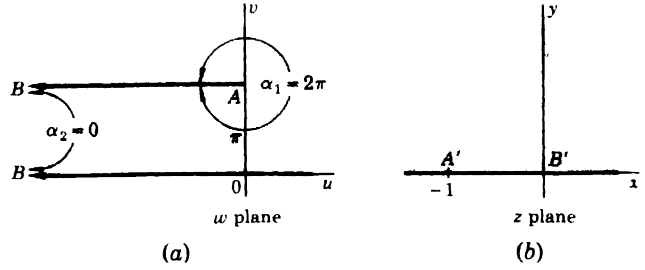


FIGURE 17.13
A semi-infinite channel to be mapped onto a half plane

and

$$(7) \qquad w = K(z + \ln z) + C$$

To determine the constants K and C, we write (7) in the form

$$u + iv = (K_1 + iK_2)(x + iy + \ln |z| + i \arg z) + C_1 + iC_2$$

from which, by equating imaginary parts, we obtain

$$(8) \qquad v = K_1 y + K_2 x + K_2 \ln |z| + K_1 \arg z + C_2$$

Now, when w becomes infinite along AB, on which $v = \pi$, the image point z approaches zero along the negative real axis, on which $y = 0$ and $\arg z = \pi$. Hence, from (8),

$$\pi = \lim_{z \to 0^-} (K_1 \cdot 0 + K_2 x + K_2 \ln |z| + K_1 \pi + C_2)$$

Obviously K_2 must be zero to keep $\ln |z|$ from making the right member infinite. Hence,

$$(9) \qquad \pi = K_1 \pi + C_2$$

Also, as w becomes infinite along OB, on which $v = 0$, the image point z approaches zero along the positive real axis, on which $y = 0$ and $\arg z = 0$ Hence, using (8) again, we have

$$0 = \lim_{z \to 0^+} (K_1 \cdot 0 + C_2) = C_2$$

Therefore, $C_2 = 0$, and so from (9) we find that $K_1 = 1$. Thus (7) reduces to

$$w = z + \ln z + C_1$$

Finally, the point $w = i\pi$ must map into the point $z = -1$ Hence, from the last equation,

$$\begin{aligned} i\pi &= -1 + \ln(-1) + C_1 \\ &= -1 + i\pi + C_1 \end{aligned}$$

and so C_1 must equal 1 The required mapping function is, therefore,

$$w = z + \ln z + 1$$

Figure 17 14 shows the curves in the w-plane which correspond to the lines $x = 0, \frac{1}{2}, 1$ and the lines $y = \pi/4, \pi/2, 3\pi/4$ The resulting configuration can be shown to represent either the lines of equal velocity potential and the streamlines for the flow of an ideal incompressible fluid from an infinite straight channel or the lines of flux and the equipotential lines for a parallel-plate condenser

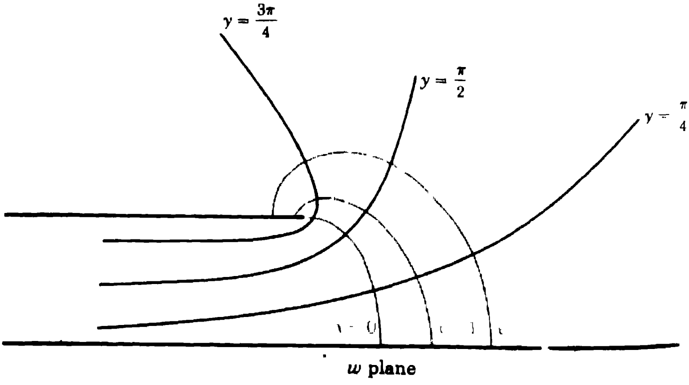
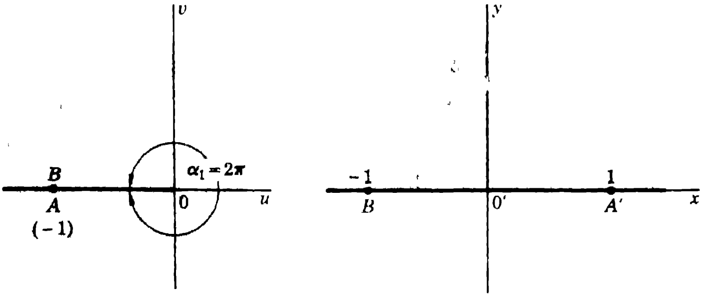


FIGURE 17.14 Typical streamlines for fluid flow from a long straight channel.

EXERCISES

1 Find the transformation which will map the region shown in Fig 17.15*a* onto the upper half plane, as indicated

FIGURE 17.15

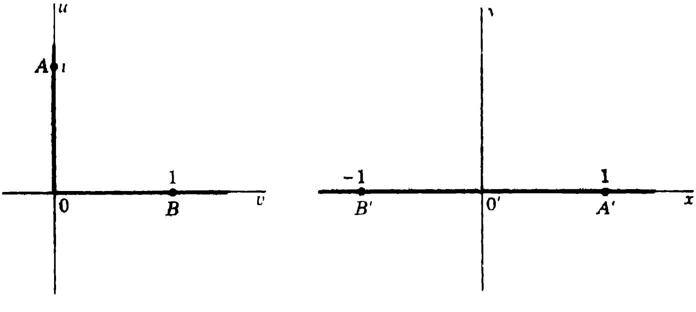


(*a*) (*b*)

2 Using the results of Exercise 1, find the steady-state temperature distribution in the w-plane if the upper side of the negative u-axis is kept at the temperature $T = 100°$ and the lower side of the negative u-axis is kept at the temperature $T = 0°$

3 Find the transformation which will map the exterior of the first quadrant in the w-plane into the upper half of the z-plane, as indicated in Fig 17.16

FIGURE 17.16

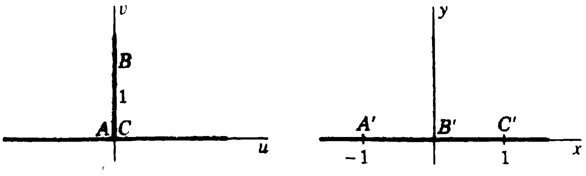


(*a*) (*b*)

4 Using the results of Exercise 3, find the equations of the isothermal curves in the exterior of the first quadrant of the w-plane if the positive u-axis is kept at the temperature $T = 100°$ and the positive half of the v-axis is kept at the temperature $T = 0°$

5 Find the transformation which will map the region shown in Fig. 17.17*a* onto the upper half plane, as indicated.

FIGURE 17.17

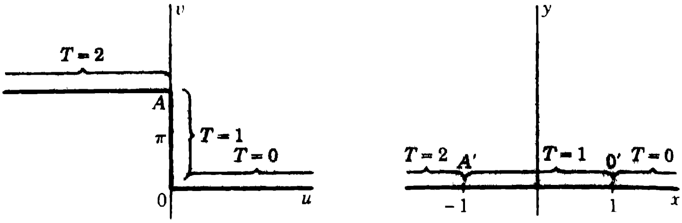


(*a*) (*b*)

6 Using the results of Exercise 5, find the steady-state temperature distribution in the upper half of the w-plane if the u-axis is kept at the temperature $T = 0°$ and the segment of the v-axis between 0 and i is kept at the temperature $T = 100°$

7 By first mapping into the z-plane as indicated, find the steady-state temperature at any point in the region shown in Fig. 17.18.

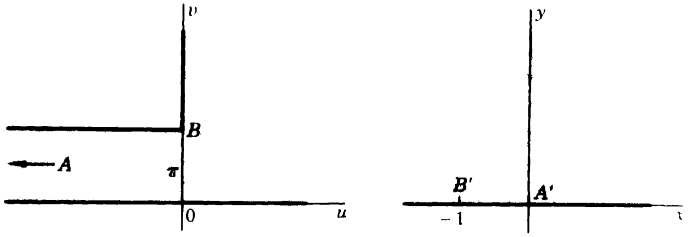
FIGURE 17.18



(a) (b)

8 Find the transformation which will map the region shown in Fig. 17.19*a* onto the upper half plane, as indicated.

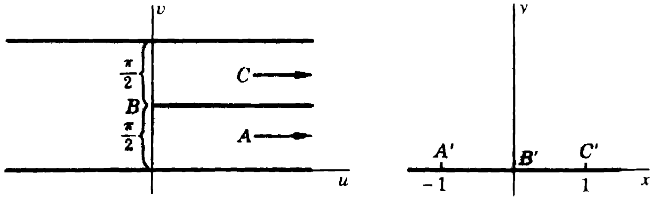
FIGURE 17.19



(a) (b)

9 Find the transformation which will map the region shown in Fig. 17.20*a* onto the upper half plane, as indicated.

FIGURE 17.20



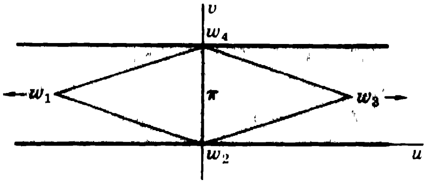
(a) (b)

10 Find the transformation which will map the interior of the infinite strip

$$0 \leq \mathscr{I}(w) \leq \pi$$

onto the upper half of the z-plane. (Hint: Consider the strip as the limiting form of the quadrilateral shown in Fig. 17.21, as w_1 and w_3 become infinite, and let w_1, w_2, and w_3 correspond, respectively, to $z = 0$, 1, and ∞, with the image of w_4 to be determined.)

FIGURE 17.21



Appendix

A.1

Graeffe's root-squaring process

At various points in our work, notably in the solution of systems of simultaneous differential equations and the determination of the inverse Laplace transforms of rational fractional functions of s, we found it necessary to solve polynomial equations of relatively high degree. Since exact formulas for the roots of a polynomial equation $P_n(x) = 0$ exist only when $n \leqq 4$ and since, when $n = 3$ and $n = 4$, these are complicated and awkward to use, it is clear that the roots of $P_n(x) = 0$ will usually have to be found by some process of numerical approximation. Numerous procedures are available for this purpose, including *Newton's method* and the *method of interpolation*. However, in many respects the best method is what is known as **Graeffe's root-squaring process,*** which has the desirable feature that it yields all the roots, both real and complex, at essentially the same time. To present the theory of this method, let us assume first that the roots of the given equation

(1) $$x^n + a_1x^{n-1} + a_2x^{n-2} + a_3x^{n-3} + \cdots + a_{n-1}x + a_n = 0$$

say $-r_1, -r_2, -r_3, \ldots, -r_n$, are all real and distinct and have been arranged in decreasing order of absolute value from $-r_1$ to $-r_n$.

Now let us rewrite Eq. (1) with all even powers of x on one side and all odd powers of x on the other

$$x^n + a_2x^{n-2} + a_4x^{n-4} + \cdots = -(a_1x^{n-1} + a_3x^{n-3} + a_5x^{n-5} + \cdots)$$

and then square both sides

---

* Named for the Swiss mathematician C. H. Graeffe (1799–1873), who published this method in 1837 in a paper which won a prize offered by the Academy of Sciences of Berlin for a practical method of computing complex roots.

$$\begin{array}{ll}
x^{2n} + a_2^2x^{2n-4} + a_4^2x^{2n-8} + & a_1^2x^{2n-2} + a_3^2x^{2n-6} + a_5^2x^{2n-10} + \\
\quad +2a_2x^{2n-2} + 2a_4x^{2n-4} + \cdots & \quad +2a_1a_3x^{2n-4} + 2a_1a_5x^{2n-6} + \cdot \\
\qquad +2a_2a_4x^{2n-6} + \cdots & \qquad + 2a_3a_5x^{2n-8} + \cdot \\
\qquad + \cdots\cdots\cdots = & \qquad + \cdots
\end{array}$$

Collecting terms again on the left, we obtain

$$x^{2n} - (a_1^2 - 2a_2)x^{2n-2} + (a_2^2 - 2a_1a_3 + 2a_4)x^{2n-4} \\ - (a_3^2 - 2a_2a_4 + 2a_1a_5 - 2a_6)x^{2n-6} + \cdot = 0 \tag{2}$$

Since Eq (2) was obtained from (1) by squaring, it is evident that it will vanish for any value of x for which Eq. (1) vanishes In other words, any root of the original equation is also a root of the derived equation (2) Now, in (2) let $y = -x^2$ Then

$$\begin{aligned}
x^{2n} &= (-y)^n = (-1)^n y^n \\
x^{2n-2} &= (-y)^{n-1} = -(-1)^n y^{n-1} \\
x^{2n-4} &= (-y)^{n-2} = (-1)^n y^{n-2} \\
\cdot\ \cdot &= \cdot \qquad =\ \cdot
\end{aligned}$$

and dividing out $(-1)^n$, Eq (2) becomes

$$y^n + (a_1^2 - 2a_2)y^{n-1} + (a_2^2 - 2a_1a_3 + 2a_4)y^{n-2} \\ + (a_3^2 - 2a_2a_4 + 2a_1a_5 - 2a_6)y^{n-3} + \cdot = 0 \tag{3}$$

By virtue of the substitution $y = -x^2$, it is evident that the roots of (3) are $-(-r_1)^2$, $-(-r_2)^2$, . . , $-(-r_n)^2$, or $-r_1^2$, $-r_2^2$, , $-r_n^2$ We have thus constructed a new equation whose roots are numerically equal to the squares of the roots of the original equation Obviously, by repeating this process, equations can be obtained whose roots are numerically equal to the fourth, eighth, sixteenth, thirty-second, . powers of the roots of Eq. (1).

The effect of this root-squaring process is to give equations whose roots are more and more widely separated For instance, if two roots of the given equation are in the ratio 5 4, then their 128th powers are in the ratio

$$5^{128} \cdot 4^{128} \quad \text{or} \quad 2\ 54 \times 10^{12}\ 1$$

This is a highly desirable situation, for, as we shall soon see, equations whose roots are widely separated can readily be solved with considerable accuracy

The squaring process which leads to equations whose roots are high powers of the roots of a given equation may be carried out systematically in tabular form as follows Write down the successive coefficients in the original equation, taking care to write zero for the coefficient of any missing term. Then under each coefficient write its square and twice all products of coefficients symmetrically located on each side of it, the signs of these products being alternately negative and positive as coefficients farther and farther from the one in question are multiplied. The

1	a_1	a_2	a_3	·	a_{n-2}	a_{n-1}	a_n
1	a_1^2 $-2a_2$	a_2^2 $-2a_1a_3$ $+2a_4$	a_3^2 $-2a_2a_4$ $+2a_1a_5$ $-2a_6$	· · · · · · ·	a_{n-2}^2 $-2a_{n-3}a_{n-1}$ $+2a_{n-4}a_n$	a_{n-1}^2 $-2a_{n-2}a_n$	a_n^2

sum of the terms placed below any coefficient in this array is the coefficient of the corresponding term in the new equation.

Now suppose that by a number of repetitions of this process an equation

(4) $$z^n + \alpha_1 z^{n-1} + \alpha_2 z^{n-2} + \cdots + \alpha_{n-1}z + \alpha_n = 0$$

has been obtained whose roots are numerically the mth powers of those of the given equation, or specifically $-r_1^m$, $-r_2^m$, $-r_3^m$, . . . , $-r_n^m$ From the usual relations between the roots and the coefficients of a polynomial equation, it follows that

$$\alpha_1 = -(\text{sum of the roots}) = r_1^m + r_2^m + r_3^m + \cdots + r_n^m$$

$$\alpha_2 = \text{sum of the roots taken two at a time}$$
$$= r_1^m r_2^m + r_1^m r_3^m + \cdots + r_2^m r_3^m + \cdots + r_{n-1}^m r_n^m$$

(5) $$= -(\text{sum of the roots taken three at a time})$$
$$= r_1^m r_2^m r_3^m + r_1^m r_2^m r_4^m + \cdots + r_2^m r_3^m r_4^m + \cdots + r_{n-2}^m r_{n-1}^m r_n^m$$
$$\cdots$$
$$\alpha_n = (-1)^n(\text{product of all the roots}) = r_1^m r_2^m r_3^m \cdots r_n^m$$

Under the hypothesis that $|r_1| > |r_2| > |r_3| > \cdots > |r_n|$ and that m is large, say 128 or 256, it follows that r_1^m is enormously larger than r_2^m, which in turn is enormously larger than r_3^m, and so on Hence, to a high degree of approximation,

$$\alpha_1 \doteq r_1^m$$
$$\alpha_2 = r_1^m r_2^m$$
$$\alpha_3 = r_1^m r_2^m r_3^m$$
$$\cdots$$
$$\alpha_{n-1} = r_1^m r_2^m r_3^m \cdots r_{n-1}^m$$
$$\alpha_n = r_1^m r_2^m r_3^m \cdots r_{n-1}^m r_n^m$$

since in each case the first term is very much larger than any or all of the succeeding terms.

Thus, $|r_1|$ can be found simply by extracting the mth root of the second coefficient in the final equation (4). Similarly, since the ratio of the third coefficient to the second is

$$\frac{\alpha_2}{\alpha_1} \doteq \frac{r_1^m r_2^m}{r_1^m} = r_2^m$$

$|r_2|$ can be found by extracting the mth root of this ratio In the same way, $|r_3|$ is the mth root of the ratio of the fourth coefficient to the third, and so on. The signs of the roots cannot be determined by this procedure, but can easily be inferred from a rough sketch of the graph of the left member of the original equation (1) supplemented by Descartes's rule of signs. Example 1 will make the details clear.

Just how many root-squaring operations must be performed in a given case, i.e., just how large m should be, cannot be told in advance. An adequate working rule is to continue until each new coefficient (with certain exceptions in the case of multiple roots and complex roots) is essentially the square of the preceding one, i.e , until the product terms in the calculation of the new coefficients make no appreciable contribution.

The case of equal roots can be handled without difficulty by returning to Eq. (4) and supposing two of the r's to be equal, say $r_3 = r_4$ Then when all but the dominant terms in each coefficient are rejected, we find from (5) that

$$\begin{aligned}
\alpha_1 &= r_1^m \\
\alpha_2 &= r_1^m r_2^m \\
\alpha_3 &= r_1^m r_2^m r_3^m + r_1^m r_2^m r_4^m = 2r_1^m r_2^m r_3^m \\
\alpha_4 &= r_1^m r_2^m r_3^m r_4^m = r_1^m r_2^m r_3^{2m} \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
\alpha_{n-1} &= r_1^m r_2^m r_3^m r_4^m r_5^m \cdots r_{n-1}^m = r_1^m r_2^m r_3^{2m} r_5^m \cdots r_{n-1}^m \\
\alpha_n &= r_1^m r_2^m r_3^m r_4^m r_5^m \cdots r_{n-1}^m r_n^m = r_1^m r_2^m r_3^{2m} r_5^m \cdots r_{n-1}^m r_n^m
\end{aligned}$$

Hence, to a high degree of approximation, the final equation is

$$z^n + r_1^m z^{n-1} + r_1^m r_2^m z^{n-2} + 2r_1^m r_2^m r_3^m z^{n-3} + r_1^m r_2^m r_3^{2m} z^{n-4} + \cdots = 0$$

Evidently, for each pair of equal roots there will be one term (the fourth in this case) which, as the root-squaring process is repeated, does not approach the square of its previous value but instead approaches one-half this value. In other words, when two equal roots are present, there is always one coefficient for which in the root-squaring procedure the product of adjacent coefficients never becomes negligible but always makes a contribution approaching one-half the square of the coefficient itself

		$r_1^m r_2^m$	$2r_1^m r_2^m r_3^m$	$r_1^m r_2^m r_3^{2m}$		
			$4r_1^{2m} r_2^{2m} r_3^{2m}$ $-2(r_1^m r_2^m)(r_1^m r_2^m r_3^{2m})$			
			$2r_1^{2m} r_2^{2m} r_3^{2m}$			

When one or more coefficients behave in this manner as successive equations are constructed, the presence of double roots is *always* indicated, and the process can be terminated as soon as all but the exceptional coefficient or coefficients are uninfluenced by the product terms.

The determination of the roots in this case, once it is recognized, is simple enough All roots except the repeated one can be found just as before by extracting the mth root of the ratios of successive pairs of nonexceptional coefficients. The repeated root can be found by extracting the $2m$th root of the ratio of the coefficients immediately following and immediately preceding the exceptional coefficient. Here again, the signs of the roots must be determined by subsequent inspection.

When the given equation contains a pair of complex roots, necessarily conjugates of each other, the analysis is a little different. To investigate this case, suppose specifically that the equation to be solved is of the fourth degree with roots $-r_1$, $-r_2e^{i\theta}$, $-r_2e^{-i\theta}$, $-r_3$, whose absolute values are such that $|r_1| > |r_2| > |r_3|$. The equation can then be written

$$(x + r_1)(x + r_2e^{i\theta})(x + r_2e^{-i\theta})(x + r_3) = 0$$

After m root-squaring operations have been performed, the resultant equation has roots

$$-r_1^m \qquad -r_2^m e^{im\theta} \qquad -r_2^m e^{-im\theta} \qquad -r_3^m$$

and can, therefore, be written

$$(z + r_1^m)(z + r_2^m e^{im\theta})(z + r_2^m e^{-im\theta})(z + r_3^m) = 0$$

or

$$\begin{aligned} z^4 &+ (r_1^m + r_2^m e^{im\theta} + r_2^m e^{-im\theta} + r_3^m)z^3 \\ &+ (r_1^m r_2^m e^{im\theta} + r_1^m r_2^m e^{-im\theta} + r_1^m r_3^m + r_2^m e^{im\theta} r_2^m e^{-im\theta} \\ &\qquad + r_2^m e^{im\theta} r_3^m + r_2^m e^{-im\theta} r_3^m)z^2 \\ &+ (r_1^m r_2^m e^{im\theta} r_2^m e^{-im\theta} + r_1^m r_2^m e^{im\theta} r_3^m + r_1^m r_2^m e^{-im\theta} r_3^m \\ &\qquad + r_2^m e^{im\theta} r_2^m e^{-im\theta} r_3^m)z \\ &+ (r_1^m r_2^m e^{im\theta} r_2^m e^{-im\theta} r_3^m) = 0 \end{aligned} \tag{6}$$

where the coefficients have been expressed at length as the appropriate symmetric functions of the roots.

In every coefficient in (6) except the coefficient of z^2, the first term is obviously the term of greatest absolute value. This is not the case in the coefficient of z^2, however, for by combining terms this can be written

$$2r_1^m r_2^m \cos m\theta + r_1^m r_3^m + r_2^{2m} + 2r_2^m r_3^m \cos m\theta$$

and it is clear that, if $\cos m\theta$ is approximately 1 or -1, the first term is dominant, whereas, if $\cos m\theta$ is approximately 0, one of the later terms is dominant. Thus, as m increases, the coefficient of z^2 continuously fluctuates in sign and does not become and remain positive, as it does when all the roots are real. This is the characteristic which identifies the presence of complex roots in polynomial equations of all degrees, as many coefficients behaving in this manner as there are pairs of complex roots.

Once the existence of complex roots is recognized, it is a simple matter to obtain the absolute values of the real roots by extracting the mth root of the ratios of successive nonexceptional coefficients, just as before. Moreover, the modulus of the complex roots can be found by taking the $2m$th root of the quotient of the coefficient which immediately follows the exceptional one divided by the coefficient which immediately precedes the exceptional one.

To complete the determination of the complex roots, for which at this stage only the absolute value is known, let those roots now be written in the form $u \pm iv$. In the original equation, the coefficient of x^{n-1} is the negative of the sum of the roots; hence,

$$a_1 = -[-r_1 + (u + iv) + (u - iv) - r_3 - \cdot \cdot \cdot]$$

and thus

$$u = \frac{-a_1 - (-r_1 - r_3 - \cdot \cdot \cdot)}{2}$$

$$= -\tfrac{1}{2}(\text{coefficient of } x^{n-1} + \text{sum of all real roots})$$

As soon as u is determined, v can be found from the familiar identity $r_2^2 = u^2 + v^2$

As we have already remarked, the presence of more than one pair of complex roots is indicated by the presence of more than one coefficient which fluctuates in sign as the root squaring continues. In this case all real roots can be found, just as before, from adjacent pairs of nonexceptional coefficients by extracting the mth root of their quotients. The moduli of the various complex roots can also be found, as before, by taking the $2m$th root of the ratios of the coefficients immediately after and immediately before each exceptional one, the exceptional coefficients being necessarily nonadjacent if the pairs of complex roots are of different absolute value The only modification required in this case is in the determination of the real parts of the various complex roots.

To illustrate this modification, let the given equation contain two pairs of complex roots

$$u_1 \pm iv_1 \quad \text{and} \quad u_2 \pm iv_2$$

of absolute values r_1 and r_2, together with additional real roots $-r_3, -r_4, \ldots$. As before, the coefficient of x^{n-1} in the original equation is the negative of the sum of the roots; hence,

$$a_1 = -[(u_1 + iv_1) + (u_1 - iv_1) + (u_2 + iv_2) + (u_2 - iv_2) - r_3 - r_4 - \cdot \cdot \cdot]$$

$$= -2u_1 - 2u_2 + r_3 + r_4 + \cdot \cdot \cdot$$

or

$$u_1 + u_2 = \frac{-a_1 - (-r_3 - r_4 - \cdot \cdot \cdot)}{2}$$

$$= -\tfrac{1}{2}(\text{coefficient of } x^{n-1} + \text{sum of all real roots}) \tag{7}$$

This is one equation in the unknown real components u_1 and u_2.

To obtain a second equation in u_1 and u_2, we make use of the fact that the coefficient of x in the original equation is equal to $(-1)^{n-1}$ (sum of the roots taken $n - 1$ at a time)

Hence,

$$
\begin{aligned}
a_{n-1} = (-1)^{n-1}[&(u_1 - iv_1)(u_2 + iv_2)(u_2 - iv_2)(-r_3)(-r_4)\cdots \\
&+ (u_1 + iv_1)(u_2 + iv_2)(u_2 - iv_2)(-r_3)(-r_4)\cdots \\
&+ (u_1 + iv_1)(u_1 - iv_1)(u_2 - iv_2)(-r_3)(-r_4)\cdots \\
&+ (u_1 + iv_1)(u_1 - iv_1)(u_2 + iv_2)(-r_3)(-r_4)\cdots \\
&+ (u_1 + iv_1)(u_1 - iv_1)(u_2 + iv_2)(u_2 - iv_2)\text{ (sum of all products of the } n - 4 \text{ real roots taken } n - 5 \text{ at a time)}] \\
= (-1)^{n-1}[&(u_1 - iv_1)r_2^2\text{ (product of the } n - 4 \text{ real roots)} \\
&+ (u_1 + iv_1)r_2^2\text{ (product of the } n - 4 \text{ real roots)} \\
&+ (u_2 - iv_2)r_1^2\text{ (product of the } n - 4 \text{ real roots)} \\
&+ (u_2 + iv_2)r_1^2\text{ (product of the } n - 4 \text{ real roots)} \\
&+ r_1^2r_2^2\text{ (sum of all products of the } n - 4 \text{ real roots taken } n - 5 \text{ at a time)}] \\
= (-1)^{n-1}[&(2u_1r_2^2 + 2u_2r_1^2)\text{ (product of the } n - 4 \text{ real roots)} \\
&+ r_1^2r_2^2\text{ (sum of all products of the } n - 4 \text{ real roots taken } n - 5 \text{ at a time)}]
\end{aligned}
$$

Hence, finally,

$$
u_1r_2^2 + u_2r_1^2 = \frac{(-1)^{n-1}a_{n-1} - r_1^2r_2^2\text{ (sum of all products of the } n - 4 \text{ real roots taken } n - 5 \text{ at a time)}}{2\text{ (product of all } n - 4 \text{ real roots)}}
$$

or

$$
u_1r_2^2 + u_2r_1^2 = \frac{(-1)^{n-1}a_{n-1}}{2\text{ (product of all } n - 4 \text{ real roots)}} - \frac{r_1^2r_2^2}{2}\text{ (sum of reciprocals of all } n - 4 \text{ real roots)} \tag{8}
$$

From Eqs (7) and (8), u_1 and u_2 can be found at once. Then v_1 and v_2 can be determined from the relations

$$
r_1^2 = u_1^2 + v_1^2 \qquad \text{and} \qquad r_2^2 = u_2^2 + v_2^2
$$

When more than two pairs of complex roots are present, this procedure can be generalized by using, in addition to the relation

$$
a_1 = -\text{(sum of all the roots)}
$$

and the x-coefficient relation, other relations arising from the coefficients of x^2, x^3, ... However, these further equations in the real components u_1, u_2, u_3, u_4, . . . are nonlinear, and solving them for u_1, u_2, u_3, u_4, . . may be very difficult.

EXAMPLE 1

Find all the roots of the equation

$$P_7(x) = x^7 + x^6 - 4x^5 - 4x^4 - 2x^3 - 5x^2 - x - 1 = 0$$

The construction of the successive equations presents no difficulty and is adequately set forth in Table A.1. By the time $m = 128$, all coefficients are uninfluenced by the product terms except the fifth and the seventh, which continually fluctuate in sign. We can, therefore, terminate the root-squaring process at this stage with the assurance that there are two pairs of complex roots and three distinct real roots.

To find the magnitudes of the three real roots, we have

$$\log |r_1| = \frac{\log (7.4844 \times 10^{42})}{128} = 0.33495$$

$$|r_1| = 2.162$$

$$\log |r_2| = \frac{\log (4.1006 \times 10^{79}) - \log (7.4844 \times 10^{42})}{128} = 0.28702$$

$$|r_2| = 1.936$$

$$\log |r_3| = \frac{\log (7.2835 \times 10^{100}) - \log (4.1006 \times 10^{79})}{128} = 0.16601$$

$$|r_3| = 1.466$$

Since there is only one change of sign between successive coefficients in the original equation, only one of the three real roots can be positive. Since $P_7(2) = -39$ and $P_7(3) = 1{,}517$, the positive root must lie between 2 and 3. Hence, the real roots are approximately 2.162, −1.936, −1.466.

To find the absolute values of the complex roots, we extract the 256th root of the ratios of the coefficients just after and just before the exceptional coefficients. Thus

$$\log r_4 = \frac{\log (4.1083 \times 10^{79}) - \log (7.2835 \times 10^{100})}{256} = 9.91700 - 10$$

$$r_4 = 0.8264$$

$$\log r_5 = \frac{\log (1) - \log (4.1083 \times 10^{79})}{256} = 9.68901 - 10$$

$$r_5 = 0.4887$$

The real parts of these roots must satisfy Eqs. (7) and (8), hence

$$u_4 + u_5 = \frac{-1 - [(2.162) + (-1.936) + (-1.466)]}{2}$$

or $$u_4 + u_5 = 0.120$$

and $$u_4(0.4887)^2 + u_5(0.8264)^2 = \frac{(-1)^6(-1)}{2(2.162)(-1.936)(-1.466)} - \frac{(0.8264)^2(0.4887)^2}{2}\left(\frac{1}{2.162} + \frac{1}{-1.936} + \frac{1}{-1.466}\right)$$

or $$0.239u_4 + 0.683u_5 = -0.021$$

Solving these two equations simultaneously, we find without difficulty that

$$u_4 = 0.233 \qquad \text{and} \qquad u_5 = -0.113$$

Hence, $$v_4 = \sqrt{r_4^2 - u_4^2} = 0.795$$

$$v_5 = \sqrt{r_5^2 - u_5^2} = 0.476$$

The complex roots of the given equation are therefore approximately

$$0.233 \pm 0.795i \qquad \text{and} \qquad -0.113 \pm 0.476i$$

table A.1

	1	1	−4	−4	−2	−5	−1	−1
$m = 2$	1	9	20	−8	−26	29	−9	1
$m = 4$	1	$4\ 1000 \times 10$	$4\ 9200 \times 10^{2}$	$1\ 6440 \times 10^{3}$	$7\ 6200 \times 10^{2}$	$3\ 5700 \times 10^{2}$	$2\ 3000 \times 10$	1
$m = 8$	1	$6\ 9700 \times 10^{2}$	$1\ 0877 \times 10^{5}$	$1\ 9821 \times 10^{6}$	$-5\ 7061 \times 10^{5}$	$9\ 5686 \times 10^{4}$	$-1\ 8500 \times 10^{2}$	1
$m = 16$	1	$2\ 6827 \times 10^{6}$	$9\ 0669 \times 10^{9}$	$4\ 0530 \times 10^{12}$	$-5\ 3760 \times 10^{10}$	$8\ 9486 \times 10^{9}$	$-1\ 5715 \times 10^{5}$	1
$m = 32$	1	$5\ 3835 \times 10^{10}$	$8\ 0034 \times 10^{19}$	$1\ 6428 \times 10^{25}$	$-6\ 9647 \times 10^{22}$	$8\ 0060 \times 10^{19}$	$6\ 7989 \times 10^{9}$	1
$m = 64$	1	$2\ 7381 \times 10^{21}$	$6\ 4036 \times 10^{39}$	$2\ 6988 \times 10^{50}$	$2\ 2202 \times 10^{45}$	$6\ 4096 \times 10^{39}$	$-1\ 1390 \times 10^{20}$	1
$m = 128$	1	$7\ 4844 \times 10^{42}$	$4\ 1006 \times 10^{79}$	$7\ 2835 \times 10^{100}$	$1\ 4697 \times 10^{90}$	$4\ 1083 \times 10^{79}$	$1\ 5400 \times 10^{38}$	1

EXERCISES

Find all the roots of each of the following equations

1 $x^3 - 6x^2 + 11x - 7 = 0$

2 $x^3 + 2x^2 + 2x + 2 = 0$

3 $x^4 - x^3 - 10x^2 - x + 1 = 0$

4 $4x^4 + 16x^3 + 25x^2 + 21x + 9 = 0$

5 $16x^5 - 16x^4 - 12x^3 + 12x^2 - 1 = 0$

6 $x^5 - 5x^3 + 4x - 10 = 0$

7 $x^6 - 8x^4 + 17x^3 - 10x^2 + 10 = 0$

8 Discuss the application of the root-squaring process to an equation with a triple root.

9 Discuss the application of the root-squaring process to an equation with two pairs of complex roots of equal moduli

10 Discuss the application of the root-squaring process to an equation with a pair of complex roots and a real root whose absolute value is equal to the modulus of the complex roots

Answers to Odd-numbered Exercises

Chapter 1

sec. 1.2
p 7

1 Second-order, ordinary, nonlinear
3 Second-order, ordinary, nonlinear
5 Second-order, ordinary, linear
7 Second-order, partial, linear
19 $y''' - 2y'' - y' + 2y = 0$
21 $x^2y'' - 2xy' + 2y = 0$
23 $y'' - 4y = 0$
25 $2yy'' - (y')^2 = 0$

sec. 1.3
p 11

1 $y = cx^3$
3 $y - \ln|y + 1| - x^2 = c$
5 $y = (2 - cx)/(1 - cx)$
7 $y^2 = ce^{-x}/x$
9 $e^y(y - 1) + e^{-x} = c$
11 $y = 2x^3 + 2$
13 No Yes, $y = \begin{cases} -2x^3 + 1 & x < 0 \\ x^3 + 1 & x \geqq 0 \end{cases}$
19 $x + e^{-2x-y+1} = c$

sec. 1.4
p 14

1 When the coefficient of dx is simpler than the coefficient of dy
3 $y = x + ce^{2x/(y-x)}$
5 $y^2 = cx^4 - x^2$
7 $y^3 = x^3(8 - \ln|x|)$
9 $(y - 2x)^2(y + x) = 27$
11 $y^3 = x^3 - \frac{1}{4}(7\sqrt{2}\,x^{3/2})$
13 $y^2 = \left(\dfrac{\ln|ex| - 1}{\ln|ex|}\right)x^2$
17 $(y - 3)^2 + 2(x + 2)(y - 3) - (x + 2)^2 = c$
19 $\sqrt{x^2 + y^2}$

sec. 1.5
p 18

1 $x^3 - y^2 - 3x^2y = c$
3 $\frac{1}{4}x^4 + \frac{1}{3}y^3 + x^2y = c$
5 $2cy = 2x + x^2y + 4y\ln|y|$
7 $\ln|cxy| = -1/xy$
9 $x^2y^3 = (x/y) + c$
11 $x^2 - y^2 + 2xy = c$ The equation is both homogeneous and exact
13 $x^2y^2 - x^2 - y^2 = c$ The equation is both separable and exact

sec. 1.6
p 21

1 $y = \dfrac{1}{2} - \dfrac{1}{x} + \dfrac{c}{x^2}$ The equation is also exact.
3 $y = e^{-x}\left(1 + \dfrac{c}{x}\right)$
5 $y = (1 - x)\left(c - \dfrac{x^2}{2}\right)$
7 $y = \dfrac{2\sin^2 x}{3} + \dfrac{1}{3\sin x}$

9 $y = \frac{4}{5}x^2 + cx^{-1/2}$ The equation is also homogeneous

11 $x = \frac{4}{5}y^2 + cy^{-3}$

13 $y = x + \sqrt{1 + x^2}$

15 $y = \begin{cases} x^2 & x < 0 \\ x^2 - 2x^3 & x \geqq 0 \end{cases}$

17 $y^{-2} = x^2 + cx$

19 $y^2 = x - \frac{1}{2} + ce^{-2x}$

sec. 1.7
p 25

1 $p = 14.7e^{-0.0000385h}$

3 4.27 per cent, 2.91 per cent

5 2.4 times

7 $Q = 30 - 1500/(t + 50)$

9 242 days

11 $\omega = \omega_0 e^{-kt/I}$ The flywheel will never come to rest!

13 $v = (w/k)(1 - e^{-kgt/w})$, $s = (w/k)[t - (w/kg)(1 - e^{-kgt/w})]$

15 $v = \sqrt{\frac{2k}{m}}\sqrt{\frac{1}{y} - \frac{1}{y_0}}$ $\sqrt{\frac{2k}{my_0}}\,t = \frac{y_0}{2}\cos^{-1}\frac{2y - y_0}{y_0} + \sqrt{y_0 y - y^2}$

17 $v = -x_0 \sin\sqrt{k/m}\,t$ $x = x_0 \cos\sqrt{k/m}\,t$

19 $Q = 2(100 - t) - 150[(100 - t)/100]^3$ $0 \leqq t \leqq 100$

21 $Q = 20 - 960/(t + 48)$

23 $T = 20 + 80e^{-0.0693t}$

25 $Q = (T_0 - T_1)k/h$, where k is the thermal conductivity of the material of the wall, $T = T_0 - (T_0 - T_1)x/h$

27 $Q = \frac{4k\pi r_0 r_1(T_0 - T_1)}{r_1 - r_0}$, where k is the thermal conductivity of the material of the sphere, $T = \frac{r_1T_1 - r_0T_0}{r_1 - r_0} - \frac{r_0r_1(T_1 - T_0)}{r_1 - r_0}\frac{1}{r}$

29 $i = (E/R)(1 - e^{-Rt/L})$, $0.693L/R$

31 $i = \frac{E_0}{R^2 + \omega^2L^2}(R\cos\omega t + L\sin\omega t - Re^{-Rt/L})$, $\delta = \tan^{-1}\frac{\omega L}{R}$

33 $y = r_0 e^{\rho\pi r_0^2 x/2W}$

35 $\frac{3r^2h\sqrt{g}}{R^2}t = (2h)^{3/2} - (2y)^{3/2} - h^{3/2} + (2y - h)^{3/2}$ $y \geqq \frac{h}{2}$

$\frac{r^2\sqrt{2g}}{R^2}(t - t_{h/2}) = \sqrt{2h} - 2\sqrt{y}$ $y < \frac{h}{2}$

where $t_{h/2}$ is the time it takes the tank to drain to a depth of $h/2$. The time for the tank to drain completely is

$$t_0 = \frac{R^2}{r^2}\,\frac{2^{3/2} + 1}{3}\sqrt{\frac{h}{g}}$$

37 $\frac{1}{\sqrt{y}} - \frac{1}{\sqrt{h}} = \frac{\sqrt{2g}}{3\pi r^2}wt$ The tank will never be completely empty!

39 $x^2 = -y^2 \ln|cy|$

Chapter 2

sec. 2.1
p 35

1 a $y = c_1 \sin x + c_2 \cos x$

b $y = c_1e^{-x} + c_2e^{-3x}$

c $y = c_1e^x + c_2xe^{-x}$

d $y = c_1(x - 1) + c_2(-x^2 + x - 1)$

e $y = c_1x + c_2x^{-4}$

f $y = c_1x + c_2xe^x$

sec. 2.2
p. 41

1 Dy is a function, namely, the derivative of y, whereas yD is merely an operator

3 $(D + x)(D + 2x) = (2x^2 + 3x + 3)e^x$ $(D + 2x)(D + x) = (2x^2 + 3x + 2)e^x$ These expressions differ because, in permuting the operational coefficients, variable terms are moved across symbols of differentiation

5 $y = c_1e^{x} + c_2e^{-2x}$ **7** $y = c_1e^{-\sqrt{5}x} + c_2e^{\sqrt{5}x}$

9 $y = c_1e^{-x/2} + c_2xe^{x/2}$ **11** $y = e^{-3x/10}\left(A\cos\frac{x}{10} + B\sin\frac{x}{10}\right)$

13 $5y = 6e^{-4x} + 14e^{x}$ **15** $4y = e^{2x} + 3e^{-2x}$

17 $y = 3xe^{-3x}$

19 There is no solution satisfying the given conditions except $y \equiv 0$

25 Yes

sec. 2.3
p 49

1 $y = c_1e^{-x} + c_2e^{-3x} + (3x - 7)/9$

3 $y = c_1 + c_2e^{-x} + x + x^2/2$

5 $y = e^{-x}(A\cos 3x + B\sin 3x) + (75x^2 - 30x - 9)/250$

7 $y = c_1e^{x} + c_2e^{-x} + \frac{1}{2}xe^{x} + \frac{2}{3}e^{2x}$

9 $y = c_1\cos x + c_2\sin x + \frac{1}{5}e^{x}(-2\cos x + \sin x)$

11 $y = c_1e^{-x} + c_2xe^{-x} + \frac{1}{2} - \frac{1}{50}(3\cos 2x - 4\sin 2x)$

13 $A = 1$ **15** $y = e^{-2x}(6\sin x - 2\cos x) + 2e^{x}$

17 In the limit when $\omega \to k$, Y becomes $-(t\cos kt)/2k$, which is the particular integral that would have been obtained by applying the methods of this section to the equation $y'' + k^2y = \sin kt$

sec. 2.4
p 51

1 $y = c_1e^{-2x} + c_2xe^{-2x} - e^{-2x}\ln|x|$

3 $y = c_1e^{-x} + c_2xe^{-x} + \frac{1}{4}x^2(2\ln|x| - 3)e^{-x}$

5 $Y = \frac{1}{2} - \frac{x}{4} + \frac{x}{2} - \frac{1}{2x}\ln|x + 1|$

7 $Y = -\frac{1}{2}x\ln^2|x| - x\ln|x| - x$

9 $y = c_1e^{-(a+b)x} + c_2e^{-(a-b)x} + \frac{1}{2b}\int_0^x [e^{-(a-b)(x-s)} - e^{-(a+b)(x-s)}]f(s)\,ds$

11 $y = c_1e^{-ax} + c_2xe^{-ax} + \int_0^x (x - s)e^{-a(x-s)}f(s)\,ds$

sec. 2.5
p 55

1 $y = c_1e^{x} + c_2e^{-2x} + c_3e^{3x} + x - 3$

3 $y = c_1e^{2x} + e^{2x}(c_2\cos x + c_3\sin x) + 3\cos x - \sin x$

5 $y = c_1e^{x} + c_2e^{-x} + c_3\cos 3x + c_4\sin 3x - \frac{9x^2 + 16}{81} - \frac{\sin 2x}{25}$

7 $y = c_1 + c_2e^{x} + e^{-x/2}[A\cos(\sqrt{3}\,x/2) + B\sin(\sqrt{3}\,x/2)] - x^3/3$

9 $y = \frac{4e^{2x} - 15e^{-x} + 5e^{x}}{60} + \frac{\cos x - 2\sin x}{10}$

11 $y = \frac{1}{5}(2e^{2x} + 3\cos x + \sin x)$

13 $Y = \frac{1}{2}\int_0^x (e^{x-s} - 2e^{2(x-s)} + e^{3(x-s)})f(s)\,ds$

sec. 2.6
p 62

1 $y = c_1x + c_2x\ln|x| + c_3/x$ **3** $y = \frac{c_1}{x} + \frac{c_2}{\sqrt{x}} + \frac{x}{2} + 2$

7 $2\pi\sqrt{hw/\rho g}$, where ρ is the density of water

9 $r = a\cosh \omega t$

11 In each case the deflection is equal to $\begin{cases} \dfrac{x_0^2(x_0 - 3x_1)}{6EI} & x_1 \geqq x_0 \\ \dfrac{x_1^2(x_1 - 3x_0)}{6EI} & x_1 \leqq x_0 \end{cases}$

13 $y_n = \sin\frac{(2n + 1)\pi x}{2L}$ $F_n = \frac{(2n + 1)^2\pi^2EI}{4L^2}$

15 The critical speeds and the associated deflection curves are

$$\omega_n = \frac{z_n^2 EIg}{A\rho L^2}$$

and $$y_n = (\cos z_n - \cosh z_n)\left(\cos z_n \frac{x}{l} - \cosh z_n \frac{x}{l}\right) + (\sin z_n + \sinh z_n)\left(\sin z_n \frac{x}{l} - \sinh z_n \frac{x}{l}\right)$$

where z_n is the nth one of the roots of the equation

$$\cos z \cosh z = 1$$

17 $\omega_n = \dfrac{1}{2\pi}\sqrt{\dfrac{kr^2 g}{WR^2 + Ig}}$

19 $\omega_n = \dfrac{1}{2\pi}\sqrt{\dfrac{3kL^2 g}{3Wl^2 + wL^2}}$

21 $y = a \cosh \sqrt{g/L}\, t$

23 $\omega_n = \dfrac{1}{2\pi}\sqrt{\dfrac{3(WL + kl^2)g}{3(W + w)L^2}}$

25 (a) $\omega_n = \dfrac{1}{2\pi}\sqrt{\dfrac{2k}{I}}$ (b) $\omega_n = \dfrac{1}{2\pi}\sqrt{\dfrac{k(I_1 + I_2)}{I_1 I_2}}$

Chapter 3

sec. 3.2
p 72

1 $x = 4,\ y = \frac{1}{3}e^{-t} - 5$

3 $x = c_1e^t + c_2e^{-2t} - \frac{9}{2}$
$y = \frac{1}{2}(-3c_1e^t - 6c_2e^{-2t} + e^{-t} + 15)$

5 $x = c_1e^{-t} + c_2e^{-2t} - (14 \cos 2t + 23 \sin 2t)/29$
$y = \frac{1}{4}[2c_1e^{-t} + c_2e^{-2t} - (88 \cos 2t + 104 \sin 2t)/29]$

7 $x = e^{-t}(A \cos t + B \sin t)$
$y = -\frac{1}{2}e^{-t}[(A + B) \cos t - (A - B) \sin t]$

9 $x = c_1 \cos t + c_2 \sin t + c_3 \cos 2t + c_4 \sin 2t + \frac{23}{10} \cos 3t$
$y = \frac{1}{2}(2c_1 \cos t + 2c_2 \sin t - 7c_3 \cos 2t - 7c_4 \sin 2t - \frac{73}{5} \cos 3t)$

11 $x = (c_1 + 1)e^{-t} + c_2e^{-4t} + 11c_3e^{4t}$
$y = -2c_1e^{-t} + c_2e^{-4t} + 3c_3e^{4t} + 1$
$z = c_1e^{-t} + c_2e^{-4t} - 29c_3e^{4t}$

13 $x = e^t$, $y = e^t$, and $z = e^t$

15 $x = c_1e^{-t} + c_2e^t + \int_0^t \sinh (x - s)[-z''(s) - z'(s) + z(s)]\, ds$
$y = c_1e^{-t} - c_2e^t + \int_0^t \sinh (x - s)[-z''(s) + 2z(s)]\, ds$
$z = z$ (z arbitrary)

17 $(5D - 4)x - (4D - 5)y = 0 \qquad (D^2 - 2D)x + (D - 2)y = 0$

19 $Q_1 = 100(1 - e^{-t/10}),\ Q_2 = 100(1 + e^{-t/10})$

sec. 3.3
p. 78

1 $x = c_1e^{-3t} - \frac{1}{2},\ y = c_1e^{-3t} + \frac{1}{2}$

3 $x = c_1e^t + 2c_2e^{2t} + t,\ y = -c_1e^t + c_2e^{2t} + 1$

5 $x = -2c_1 + c_2e^{-5t} + \frac{5}{6}e^t + \frac{1}{4}e^{-t}$
$y = c_1 - 3c_2e^{-5t} - \frac{1}{2}e^t + \frac{1}{4}e^{-t}$

7 $x = -2c_1 \cos t - 2c_2 \sin t + c_3 \cos 2t + c_4 \sin 2t + \frac{1}{4}(5 - 3t)$
$y = 3c_1 \cos t + 3c_2 \sin t - 3c_3 \cos 2t - 3c_4 \sin 2t - \frac{1}{4}(7 - 5t)$

9 $x = 8c_1 \cos 2t + 8c_2 \sin 2t + c_3 \cos 3t + c_4 \sin 3t + \frac{25}{73} \cos t$
$y = -7c_1 \cos 2t - 7c_2 \sin 2t - c_3 \cos 3t - c_4 \sin 3t - \frac{29}{73} \sin t$

Chapter 4

sec. 4.1
p 89

3 (a) $P(x) = 1 + 3(x)^{(2)} + (x)^{(3)}$. The difference table can be constructed by addition from the leading differences

$P(0) = 1 \qquad \Delta P(0) = 0 \qquad \Delta^2 P(0) = 6 \qquad \Delta^3 P(0) = 6$

(b) $P(x) = -2(x) + (x)^{(2)} + 4(x)^{(3)} + (x)^{(4)}$ The difference table can be constructed by addition from the leading differences

$P(0) = 0 \qquad \Delta P(0) = -2 \qquad \Delta^2 P(0) = 2 \qquad \Delta^3 P(0) = 24$
$\Delta^4 P(0) = 24$

(c) $P(x) = 6 + 2(x) + 13(x)^{(2)} + 17(x)^{(3)} + 8(x)^{(4)} + (x)^{(5)}$ The difference table can be constructed by addition from the leading differences

$P(0) = 6 \qquad \Delta P(0) = 2 \qquad \Delta^2 P(0) = 26 \qquad \Delta^3 P(0) = 102$
$\Delta^4 P(0) = 192 \qquad \Delta^5 P(0) = 120$

9 (a) $F(x) = (x)^{-(2)} - 2(x)^{-(3)}$, (b) $F(x) = (x)^{-(1)} - 2(x)^{-(2)}$,
(c) $F(x) = (x)^{-(1)} - 4(x)^{-(2)} + 4(x)^{-(3)}$

13 $$f(x_0, x_1, \ldots, x_n) = \sum_{j=1}^{n} \left[\frac{f(x_j)}{\prod\limits_{\substack{i=0 \\ i \neq j}}^{n} (x_j - x_i)} \right]$$

sec. 4.2
p 97

3 (a) 1.338, (b) 2.819 5 $-x^3 + 5x^2 + x - 2$

9 $$x_{\max} = \frac{x_0 + x_1}{2} - \frac{f(x_0, x_1)}{2f(x_0, x_1, x_2)}$$

$$y_{\max} = y_1 - \frac{[f(x_0, x_1) - f(x_0, x_2) + f(x_1, x_2)]^2}{4f(x_0, x_1, x_2)}$$

sec. 4.3
p 106

1 $f'(200) = 0.00500000$ exact value $= 0.00500000$
$f''(200) = -0.00002499$ exact value $= -0.00002500$
$f'''(200) = 0.00000024$ exact value $= 0.00000025$
$f'(205) = 0.00487806$ exact value $= 0.00487805$
$f''(205) = -0.00002377$ exact value $= -0.00002380$
$f'''(205) = 0.00000025$ exact value $= 0.00000023$

3 $\frac{1}{4}(n + 1)^2 n^2$ 5 7.486

7

x	$\int_x^1 \frac{\sin x}{x}\,dx$	x	$\int_x^1 \frac{\sin x}{x}\,dx$
0.0	0.946	0.6	0.356
0.1	0.846	0.7	0.265
0.2	0.746	0.8	0.174
0.3	0.647	0.9	0.086
0.4	0.550	1.0	0.000
0.5	0.453		

9 When the first difference correction terms are taken into account the value of the integral is 0.31028 When the correction terms through the third differences are taken into account the value of the integral is 0.31027

13 $$y_0' = \frac{1}{h}\left(\delta f_0 - \frac{\delta^3 f_0}{24} + \frac{3\delta^5 f_0}{640} - \cdots\right)$$

15 (a) $c_0 = c_3 = \frac{3}{8}$; $c_1 = c_2 = \frac{9}{8}$
(b) $c_0 = c_4 = \frac{14}{45}$, $c_1 = c_3 = \frac{64}{45}$, $c_2 = \frac{24}{45}$

sec. 4.4
p 116

1 $y_6 = -1.7378,\ y_9 = -2.1503$

3 $y_1 = 1.1103;\ y_2 = 1.2428,\ y_3 = 1.3997$

5 $y_1 = 0.1050,\ y_2 = 0.2198,\ y_3 = 0.3445$
$z_1 = 0.9998,\ z_2 = 0.9986,\ z_3 = 0.9955$

7 Like Milne's method, the Adams-Bashforth method requires that the first few values of y be computed by other means, say by the Runge-Kutta method or the evaluation of a series expansion for the solution In particular, in the present problem only the values after $y(1.4)$ can be obtained by the Adams-Bashforth method

$y(1.1) = 1.0048 \qquad y(1.2) = 1.0187 \qquad y(1.3) = 1.0408$
$y(1.4) = 1.0703 \qquad y(1.5) = 1.1065 \qquad y(1.6) = 1.1488$

9
$$y_{n+1} = y_n + h\left(\frac{1{,}901}{720} y'_n - \frac{1{,}387}{360} y'_{n-1} + \frac{436}{120} y'_{n-2} - \frac{637}{360} y'_{n-3} + \frac{251}{720} y'_{n-4} - \quad\right) \quad \text{(open formula)}$$

$$y_{n+1} = y_n + h\left(\frac{251}{720} y'_{n+1} + \frac{323}{360} y'_n - \frac{44}{120} y'_{n-1} + \frac{53}{360} y'_{n-2} - \frac{19}{720} y'_{n-3} + \cdot \quad\right) \quad \text{(closed formula)}$$

11

x	y (by numerical solution)	y (from exact solution)
0.0	1.0000	1.0000
0.1	1.0052	1.0052
0.2	1.0214	1.0214
0.3	1.0499	1.0499
0.4	1.0919	1.0918
0.5	1.1488	1.1487
0.6	1.2222	1.2221
0.7	1.3138	1.3138
0.8	1.4256	1.4255
0.9	1.5597	1.5596
1.0	1.7184	1.7183

13 $\dfrac{h^5}{90} y^{\mathrm{v}}_{n-1}$

15
$$y = y_0 + \left(-y_0 + y_1 - \frac{y''_0 h^2}{3} - \frac{y''_1 h^2}{6}\right)\left(\frac{x}{h}\right) + \frac{y''_0 h^2}{2}\left(\frac{x}{h}\right)^2 - \left(\frac{y''_0 h^2 - y''_1 h^2}{6}\right)\left(\frac{x}{h}\right)^3$$

$y_2 = -y_0 + 2y_1 + y''_1 h^2$

An accompanying closed formula can be obtained by fitting a polynomial to y''_2 (which can be obtained from the given differential equation as soon as an estimate for y_2 is available) and any three of the data y_0, y''_0, y_1, y''_1

sec. 4.5
p 125

1 (a) $y = c_1(-3)^x + c_2(-4)^x$, (b) $y = (-3)^x + c_2x(-3)^x$,
(c) $y = 2^{x/2}(A \cos \frac{3}{4}\pi x + B \sin \frac{3}{4}\pi x)$, (d) $y = c_1 2^x + c_2 3^x$

3 (a) $y = c_1 3^x + c_2(-2)^x - \dfrac{1}{36}(6x + 1) + \dfrac{1}{15} x\, 3^x$

(b) $y = A \cos \dfrac{\pi x}{2} + B \sin \dfrac{\pi x}{2} + \dfrac{\sin x + \sin (x - 2)}{2(1 + \cos 2)}$

7 $$V = \begin{cases} \dfrac{\sin (n - x)\mu}{\sin n\mu} V_0 & \mu = \cos^{-1} \dfrac{2k + 1}{4} \qquad k < 3/2 \\ \dfrac{n - x}{n} V_0 & \qquad\qquad\qquad\qquad k = 3/2 \\ \dfrac{\sinh (n - x)\mu}{\sinh n\mu} V_0 & \mu = \cosh^{-1} \dfrac{2k + 1}{4} \qquad k > 3/2 \end{cases}$$

9 $$D_n = \begin{cases} 1 + n & \qquad\qquad\qquad \lambda = 2 \\ \dfrac{\sin (n + 1)\mu}{\sin \mu} & \cos \mu = \dfrac{\lambda}{2} \qquad -2 < \lambda < 2 \\ (-1)^n(1 + n) & \qquad\qquad\qquad \lambda = -2 \\ (-1)^n \dfrac{\sinh (n + 1)\mu}{\sinh \mu} & \cosh \mu = -\dfrac{\lambda}{2} \qquad \lambda < -2 \end{cases}$$

15 (a) The only solution is $y = 0$

(b) $y = \phi(x - 1)$

(c) Because $a_2 = 0$, the equation is actually of the first order, and its complete solution contains only one arbitrary constant:

$$y = A\left(-\frac{a_1}{a_0}\right)^x \qquad a_0 \neq 0$$

(d) $y = A\left(-\dfrac{a_1}{a_0}\right)^x + \Phi(x - 1)$, where $\Phi(x)$ is a particular solution of the nonhomogeneous equation $(a_0E + a_1)y = \phi(x)$

sec. 4.6
p 142

1 (a) $y = (2 + 133x)/102$, (b) $y = (-68 + 187x)/133$

3 (a) $x = 1.683$, $y = -1.847$, (b) $x = 1.739$, $y = -1.811$

5 (c) $x = \frac{1}{2}nP_{n0}(x) - \frac{1}{2}nP_{n1}(x)$

$x^2 = \frac{1}{6}(2n^2 + n)P_{n0}(x) - \frac{1}{2}n^2P_{n1}(x) + \frac{1}{6}(n^2 - n)P_{n2}(x)$

7 (a) $A = 1.000$, $a = 0.499$

(b) Using the results of part **a** as an initial approximation, linearizing via Taylor's series yields $A = 0.995$, $a = 0.498$

11 After a has been found, A may be determined by applying the method of least squares to the equations

$$y_1 = Ae^{ax_1}, \qquad y_2 = Ae^{ax_2}, \qquad \ldots, y_n = Ae^{ax_n}$$

in which A is the only unknown. This method is generally to be preferred to linearizing by taking logarithms, because it does not introduce any unwarranted weighting of the data. It is clearly preferable to both this and the use of Taylor's series on grounds of simplicity

13 $$y = \left(\frac{60}{\pi^3} - \frac{3}{\pi}\right) - \left(\frac{720}{\pi^5} - \frac{60}{\pi^3}\right)x^2 = 0.980 - 0.418x^2$$

15 $p = \bar{x} \cos \theta + \bar{y} \sin \theta$

Chapter 5

sec. 5.3
p 163

5 The first integer equal to or greater than $\dfrac{\ln 2}{2\pi} \dfrac{\sqrt{1 - (c/c_c)^2}}{c/c_c}$

13 $t_{\max} = 1/\omega_n$; $y_{\max} = v_0/\omega_n e$

17 $y = -e^{-4.2t}(2 \cos 14.4t + 7/12 \sin 14.4t)$

19 $y = -e^{-6t}(2 \cos 8t + 3/2 \sin 8t) + 2$

21 Period $= \tau = 2\pi \sqrt{a/\mu g}$, where $2a$ is the distance between the axes of the rollers. From this $\mu = 4\pi^2 a/g\tau^2$.

23 The maxima of the magnification ratio curves of Exercise 22 occur where $\omega/\omega_n = 1/\sqrt{1 - 2(c/c_c)^2}$, which is always greater than 1 for $0 < c/c_c < 1/\sqrt{2}$, that is, whenever a maximum exists

25 The amplitude of the steady-state response varies between the values

$$\frac{A_1}{\omega^2 - \omega_1^2} - \frac{A_2}{\omega^2 - \omega_2^2} \quad \text{and} \quad \frac{A_1}{\omega^2 - \omega_1^2} + \frac{A_2}{\omega^2 - \omega_2^2}$$

sec. 5.4
p 170

1 Across the resistance: $E = iR = 40(e^{-200t} - e^{-800t})$

Across the inductance $E = L\dfrac{di}{dt} = 8(-e^{-200t} + 4e^{-800t})$

Across the capacitance $E = \dfrac{1}{C}\displaystyle\int_0^t i\,dt = 8(3 - 4e^{-200t} + e^{-800t})$

3 $i = \dfrac{3{,}125t}{2}e^{-12{,}500t}$

5 $i_{ss} = 0.14 \cos (120\pi t + 19°40')$

7 $i = \dfrac{325t}{2}e^{-2{,}500t}$

9 $t = 0.00039$ sec

13 $|Z|$ is a minimum (or $1/|Z|$ is a maximum) for the undamped natural frequency $\Omega_n = 1/\sqrt{LC}$ For the magnification ratio the maximum always occurs at a frequency below the undamped natural frequency This involves no contradiction, since the magnification ratio M relates F and y, whose electrical analogues are E and Q, whereas the impedance relates E and $i = dQ/dt$

15 $|Z| = R\sqrt{1 + Q^2\left(\dfrac{\Omega}{\Omega_n} - \dfrac{\Omega_n}{\Omega}\right)^2} \qquad \delta = \operatorname{Tan}^{-1} Q\left(\dfrac{\Omega}{\Omega_n} - \dfrac{\Omega_n}{\Omega}\right)$

sec. 5.5
p 178

1 1, 2

3 1, $\sqrt{5}$

5 $x_1 = \frac{1}{3}(\cos t + 2 \cos 2t)$, $x_2 = \frac{1}{3}(\cos t - \cos 2t)$

7 $6\sqrt{6}$, $12\sqrt{5}$

9 $i_1 = -\dfrac{Q_0}{60\sqrt{LC}}\left(16 \sin \dfrac{t}{3\sqrt{LC}} + \sin \dfrac{t}{12\sqrt{LC}}\right)$

$i_2 = -\dfrac{Q_0}{20\sqrt{LC}}\left(4 \sin \dfrac{t}{3\sqrt{LC}} - \sin \dfrac{t}{12\sqrt{LC}}\right)$

11 $i_1 = E_0\sqrt{\dfrac{C}{L}}\left(\dfrac{243}{80} \sin \dfrac{3t}{\sqrt{LC}} - \dfrac{1}{240} \sin \dfrac{t}{3\sqrt{LC}} - \sin \dfrac{t}{\sqrt{LC}}\right)$

$i_2 = E_0\sqrt{\dfrac{C}{L}}\left(\dfrac{27}{10} \sin \dfrac{3t}{\sqrt{LC}} + \dfrac{1}{30} \sin \dfrac{t}{3\sqrt{LC}} - \sin \dfrac{t}{\sqrt{LC}}\right)$

13 $\omega_N = \dfrac{2}{\sqrt{LC}} \sin \dfrac{N\pi}{2n} \qquad N = 1, 2, \ldots, (n - 1)$

15 $\omega_N = 2\sqrt{\dfrac{k}{M}} \sin \dfrac{N\pi}{2(n + 1)} \qquad N = 1, 2, \ldots, n$

17 $\omega_N = 2\sqrt{\dfrac{k}{M}} \sin \dfrac{N\pi}{2n + 1} \qquad N = 1, 2, \ldots, n$

19 (a) $Q_k = \frac{E_0}{\omega}\sqrt{\frac{C}{L}}\,\frac{\sin\,(n-k+1)\mu}{\cos\frac{1}{2}(2n+1)\mu}\cos\omega t \qquad \cos\mu = 1 - \frac{CL\omega^2}{2}$

(b) $Q_k = (-1)^k\frac{E_0}{\omega}\sqrt{\frac{C}{L}}\,\frac{\sinh\,(n-k+1)\mu}{\sinh\frac{1}{2}(2n+1)\mu}\cos\omega t$

$\cosh\mu = \frac{CL\omega^2}{2} - 1$

Chapter 6

sec. 6.2
p 188

1 $a_0 = \frac{1}{2} \qquad a_n = \begin{cases} 0 & n = 2, 4, 6, \\ 1/n\pi & n = 1, 5, 9, \\ -1/n\pi & n = 3, 7, 11, \ldots \end{cases}$

$b_n = \begin{cases} 1/n\pi & n = 1, 3, 5, \ldots \\ 2/n\pi & n = 2, 6, 10, \\ 0 & n = 4, 8, 12, \end{cases}$

3 $a_n = 0 \qquad b_n = \frac{(-1)^{n+1}8n}{\pi(4n^2-1)} \qquad n = 1, 2, 3,$

5 $a_0 = 2$

$a_n = 0 \quad n = 1, 2, 3, \qquad b_n = \begin{cases} 3/n\pi & n = 1, 2, 4, 5, 7, 8, \\ 0 & n = 3, 6, 9, \end{cases}$

7 $a_n = 0 \qquad b_n = \frac{(-1)^{n+1}2}{n\pi}$

9 $a_0 = \frac{4\pi^2}{3} \qquad a_n = \frac{(-1)^{n+1}4}{n^2} \qquad n \neq 0 \qquad b_n = 0$

11 $a_n = 0 \qquad b_n = \begin{cases} -2/n\pi & n = 1, 3, 5, \\ -4/n\pi & n = 2, 6, 10, \\ 0 & n = 4, 8, 12, \ldots \end{cases}$

13 $a_0 = \frac{2}{\pi} \qquad a_1 = \frac{1}{2} \qquad a_n = \begin{cases} 0 & n \text{ odd}, n \neq 1 \\ \frac{2}{\pi(1-n^2)} & n \text{ even} \end{cases}$

$b_1 = \frac{1}{2} \qquad b_n = \begin{cases} 0 & n \text{ odd}, n \neq 1 \\ \frac{2n}{\pi(1-n^2)} & n \text{ even} \end{cases}$

sec. 6.3
p. 195

3 $a_n = \frac{2(1-e^{-1})}{1+4n^2\pi^2} \qquad b_n = \frac{4\pi n(1-e^{-1})}{1+4n^2\pi^2}$

5 $a_0 = \frac{1}{3} \qquad a_n = \frac{1}{n\pi}\sin\frac{2n\pi}{3} - \frac{3}{2n^2\pi^2}\left(1 - \cos\frac{2n\pi}{3}\right)$

$b_n = -\frac{1}{n\pi}\cos\frac{2n\pi}{3} + \frac{3}{2n^2\pi^2}\sin\frac{2n\pi}{3}$

7 $a_0 = \frac{5}{6} \qquad a_n = \begin{cases} -\frac{16}{n^3\pi^3} + \frac{4}{n^2\pi^2} & n = 1, 5, 9, \\ -\frac{16}{n^3\pi^3} & n = 2, 6, 10, \\ \frac{16}{n^3\pi^3} + \frac{4}{n^2\pi^2} & n = 3, 7, 11, \ldots \\ \frac{8}{n^2\pi^2} & n = 4, 8, 12, \ldots \end{cases} \qquad b_n = 0$

9 Yes In particular, the Fourier expansion of

$$f(t) = \begin{cases} t - t^2 & 0 \leqq t \leqq 1 \\ t - t^2 - 4t^3 - 2t^4 + k(t+1)^3t^3g(t) & -1 \leqq t \leqq 0 \end{cases}$$

where $g(t)$ is any function possessing a continuous second derivative, will converge to $t - t^2$ for $0 \leqq t \leqq 1$ and will have coefficients decreasing as $1/n^4$

11 Since $1/(2 + \cos t)$ possesses derivatives of all orders at all points in the interval $0 \leqq t \leqq 2\pi$, the Fourier coefficients of this function will decrease faster than the reciprocal of any fixed power of n (See Exercise 29, Sec 16 2)

15 a_n will decrease faster than $1/n^2$ provided that, for all values of n,

$$\sum_i [f'(t_i^+) - f'(t_i^-)] \cos \frac{n\pi t}{p}{}_i = 0$$

where $\{t_i\}$ is the set of points of discontinuity of f'
b_n will decrease faster than $1/n^2$ provided that, for all values of n,

$$\sum_i [f'(t_i^+) - f'(t_i^-)] \sin \frac{n\pi t}{p}{}_i = 0$$

where $\{t_i\}$ is the set of points of discontinuity of f'

sec. 6.4
p 200

1 $$A_n = \frac{2(1 - \cos \frac{1}{2}n\pi)}{n\pi} \qquad \gamma_n = \tan^{-1}\left(\frac{1 - \cos \frac{1}{2}n\pi}{\sin \frac{1}{2}n\pi}\right)$$
$$\delta_n = \tan^{-1}\left(\frac{\sin \frac{1}{2}n\pi}{1 - \cos \frac{1}{2}n\pi}\right)$$

3 $$A_n = \begin{cases} \dfrac{2}{n\pi}\left(1 - \cos \dfrac{n\pi}{2}\right) & n \text{ even} \\ \dfrac{2}{n\pi} \sin \dfrac{n\pi}{2} & n \text{ odd} \end{cases}$$

$$\gamma_n = \tan^{-1}\left(\frac{1 - 2\cos \frac{1}{2}n\pi + \cos n\pi}{\sin \frac{1}{2}n\pi}\right)$$
$$\delta_n = \tan^{-1}\left(\frac{\sin \frac{1}{2}n\pi}{1 - 2\cos \frac{1}{2}n\pi + \cos n\pi}\right)$$

5 $$c_n = \frac{1}{2ni\pi}(1 - e^{-ni\pi}) \qquad c_0 = \tfrac{1}{2}$$

7 $$c_n = \frac{1}{4n^2\pi^2}(e^{-2ni\pi} - 1) - \frac{e^{-2ni\pi}}{2ni\pi} \qquad c_0 = \tfrac{1}{2}$$

9 $$c_n = \frac{(-1)^n}{2\pi}\,\frac{2}{1 - 4n^2}$$

sec. 6.5
p. 205

1 The complete solution will originally appear in the form $y = c_1y_1 + c_2y_2 + Y$, where Y is the Fourier series obtained as the answer to Example 2. Imposing initial conditions of displacement and velocity will thus lead to a pair of simultaneous linear equations in c_1 and c_2 in which the constant terms will involve the infinite series which result from the evaluation of Y and Y' when $t = 0$. Although there is no theoretical problem in determining c_1 and c_2 from these equations, the arithmetical complications are obvious.

3 $y_{ss} = 0.190 \sin (2\pi t - 3.2^\circ) + 0.142 \sin (6\pi t - 22.3^\circ) + 0.047 \sin (10\pi t - 159.8^\circ) + 0.011 \sin (14\pi t - 171.2^\circ) + \cdots$

5 $y_{ss} = F_0[1/6 + 0.225 \cos (\pi t - 3.2^\circ) + 0.171 \cos (3\pi t - 22.3^\circ) + 0.056 \cos (5\pi t - 159.8^\circ) + \cdots$

7 $$i_{ss} = \sum_{n=-\infty}^{\infty} \frac{iE_0 e^{200ni\pi t}}{200[600n\pi + i(200n^2\pi^2 - 1{,}250)]}$$

(Note The term corresponding to $n = 0$ is to be omitted)

sec. 6.6 p. 210

1 $f(x) = 2.445 - 0.939 \cos x - 0.378 \cos 2x - 0.230 \cos 3x - 0.167 \cos 4x - 0.133 \cos 5x - 0.113 \cos 6x - \cdots$

3 $$\begin{aligned} T_{\max} = 59.46 &- 20.99 \cos t - 0.42 \cos 2t + 0.17 \cos 4t - 0.02 \cos 5t \\ &+ 0.06 \cos 7t - 0.08 \cos 8t + 0.01 \cos 10t \\ &- 8.94 \sin t - 0.20 \sin 2t + 0.20 \sin 3t + 0.43 \sin 4t \\ &- 0.07 \sin 5t + 0.02 \sin 7t + 0.03 \sin 9t \\ &- 0.05 \sin 10t \end{aligned}$$

$$\begin{aligned} T_{\min} = 45.08 &- 19.66 \cos t - 0.27 \cos 2t + 0.89 \cos 3t + 0.04 \cos 4t \\ &- 0.06 \cos 5t - 0.04 \cos 7t + 0.04 \cos 8t \\ &- 0.05 \cos 9t + 0.02 \cos 10t \\ &- 9.41 \sin t + 0.28 \sin 2t - 0.20 \sin 3t + 0.36 \sin 4t \\ &+ 0.18 \sin 5t - 0.17 \sin 6t - 0.25 \sin 7t \\ &- 0.07 \sin 8t - 0.03 \sin 9t + 0.09 \sin 10t \end{aligned}$$

5 If m is so large that all coefficients after b_m are negligibly small, the formula of the exercise expresses b_m in terms of selected values of the given function which either will be known or can easily be found By repeating this process for decreasing values of m, using previously computed coefficients where appropriate, the coefficients down to and including b_1 can be approximated.

sec. 6.7 p. 220

1 (a) $a_n \equiv 0 \qquad b_n = \dfrac{2}{\pi}\,\dfrac{1 - \cos \omega_n}{\omega_n}\,\Delta\omega \qquad \omega_n = \dfrac{n\pi}{p} \qquad \Delta\omega = \dfrac{\pi}{p}$

(b) $a_n = \dfrac{2}{\pi}\,\dfrac{1 - \cos \omega_n}{\omega_n^2}\,\Delta\omega \qquad b_n \equiv 0 \qquad \omega_n = \dfrac{n\pi}{p} \qquad \Delta\omega = \dfrac{\pi}{p}$

(c) $a_n = \dfrac{4}{\pi}\,\dfrac{\sin \omega_n - \omega_n \cos \omega_n}{\omega_n^3}\,\Delta\omega \qquad b_n \equiv 0 \qquad \omega_n = \dfrac{n\pi}{p} \qquad \Delta\omega = \dfrac{\pi}{p}$

(d) $a_n \equiv 0 \qquad b_n = \dfrac{2 \sin \omega_n}{\pi^2 - \omega_n^2}\,\Delta\omega \qquad \omega_n = \dfrac{n\pi}{p} \qquad \Delta\omega = \dfrac{\pi}{p}$

3 (a) $f(t) = \dfrac{2a}{\pi} \displaystyle\int_0^\infty \dfrac{\cos \omega t}{a^2 + \omega^2}\, d\omega$

(b) $f(t) = \dfrac{1}{\pi} \displaystyle\int_0^\infty \dfrac{a \cos \omega t + \omega \sin \omega t}{a^2 + \omega^2}\, d\omega$

(c) $f(t) = \dfrac{2}{\pi} \displaystyle\int_0^\infty \dfrac{\sin \pi\omega}{1 - \omega^2} \sin \omega t\, d\omega$

(d) $f(t) = \dfrac{2}{\pi} \displaystyle\int_0^\infty \dfrac{\cos \frac{1}{2}\pi\omega}{1 - \omega^2} \cos \omega t\, d\omega$

(e) $f(t) = \dfrac{1}{\pi} \displaystyle\int_0^\infty \dfrac{\sin \omega(1 - t) + \sin \omega t}{\omega}\, d\omega$

(f) $f(t) = \dfrac{4}{\pi} \displaystyle\int_0^\infty \dfrac{\sin \omega - \omega \cos \omega}{\omega^3} \cos \omega t\, d\omega$

5 $f(t) \doteq \frac{2}{\pi}\int_0^{\omega_0} \frac{1-\cos\omega}{\omega^2}\cos\omega t\,d\omega$

$= \frac{1}{\pi}[(t-1)\operatorname{Si}\omega_0(t-1) - 2t\operatorname{Si}\omega_0 t + (t+1)\operatorname{Si}\omega_0(t+1)]$

9 $Y = \frac{2}{\pi}\int_0^{\infty} \frac{1-\cos\omega}{\omega^2}\,\frac{(b-\omega^2)\cos\omega t + a\omega\sin\omega t}{(b-\omega^2)^2 + a^2\omega^2}\,d\omega$

11 $\frac{1}{\pi k}\int_0^{\infty} M(\omega)\int_{-\infty}^{\infty} f(s)\cos[\omega s - \omega t + \alpha(\omega)]\,ds\,d\omega$

where k is the modulus of the spring in the system, $M(\omega)$ is the magnification ratio, and $\alpha(\omega)$ is the phase angle

Chapter 7

sec. 7.1
p 232

5 No; for instance, the abscissa of e^{-t} is $\alpha_0 = -1$, whereas the abscissa of convergence of $\int_0^t e^{-t}\,dt$ is $\alpha_1 = 0$

sec. 7.2
p. 236

1 $\mathcal{L}\{f^{(n)}\} = s^n\mathcal{L}\{f\} - \sum_{j=0}^{n-1} s^{n-1-j}f^{(j)}(0^+)$

9 If $T(f')$ and $T(f'')$ are not to involve the evaluation of f or any of its derivatives, it is necessary that

$K(s,a) = K(s,b) = 0$

and that $\left.\frac{\partial K(s,t)}{\partial t}\right|_{t=a} = \left.\frac{\partial K(s,t)}{\partial t}\right|_{t=b} = 0$

If $\phi(s,t)$ is an arbitrary differentiable function which is bounded at $t = a$ and at $t = b$, these conditions are met by any kernel of the form

$K(s,t) = (t-a)^2(t-b)^2\phi(s,t)$

sec. 7.3
p 241

1 $\frac{s}{s^2 - b^2}$

3 $\frac{1}{2}\left(\frac{1}{s} + \frac{s}{s^2 + 4b^2}\right)$

5 (a) e^{-2t}; (b) $\frac{1}{6}t^3$, (c) $\frac{1}{3}\sin 3t$, (d) $2\cos 3t + \sin 3t$,
(e) $\frac{1}{2}(3e^{3t} - e^{-t})$

7 $y = \frac{1}{4}(-2e^{-3t} - 3e^{-t} + 5e^t)$, $z = \frac{1}{2}(e^{-3t} + 2e^{-t} - e^t)$

9 (a) $\Gamma(\frac{1}{2}) = 1.7725$; (b) 2,
(c) $\frac{1}{3} + \Gamma(\frac{5}{3}) + \Gamma(\frac{4}{3}) = 2\ 1291$; (d) $\Gamma(c+1)/(\ln c)^{c+1}$

sec. 7.4
p 253

1 $\frac{e^{-as}}{s}$

3 $2\frac{1+2s+2s^2}{s^3}e^{-2s}$

5 $\frac{e^{-(s-2)}}{s-2}$

7 $\frac{1+e^{-\pi s}}{s^2+1}$

9 $\frac{1}{s^2}e^{-2s} - \frac{s+1}{s^2}e^{-3s}$

11 $\frac{1}{2}\left(6\cot^{-1}\frac{s}{3} + s\ln\frac{s^2}{s^2+9}\right)$

13 $\frac{4(s+3)}{(s^2+6s+25)^2}$

15 $\frac{4}{(s^2+6s+13)^2}$

17 $\cot^{-1}\frac{s+3}{2}$

19 $\frac{1}{s}\cos^{-1}\frac{s+3}{2}$

21 $\frac{t^3e^{-2t}}{6}$

23 $\frac{1}{9}\left(e^{-t/3}\cos\frac{2t}{3} + e^{-t/3}\sin\frac{2t}{3}\right)$

25 $t - 1 + e^{-t}$

27 $\frac{1}{2} \sin 2(t-2)u(t-2)$

29 $\dfrac{(t-1)^2e^{-(t-1)}}{2}u(t-1)$

31 $\dfrac{e^{-bt} - e^{-at}}{t}$

33 $\dfrac{1 + e^{-t} - 2\cos t}{t}$

35 $\displaystyle\int_0^t \frac{\sin t}{t}\,dt$

37 $\frac{1}{2}te^{-2t}\sin t$

39 $\sin t - t\cos t$

41 $f(t)$, $f'(t)$, and $f''(t)$ are each piecewise regular and of exponential order; $f(t)$, $f'(t)$, $f''(t)$, and $f'''(t)$ are each piecewise regular and of exponential order

$$f^{(n)}(0^+) = \lim_{s\to 0} s[s^n\mathcal{L}\{f\} - \sum_{j=0}^{n-1} s^{n-1-j}f^{(j)}(0^+)]$$

43 $\frac{1}{4}(7e^{-t} + 2te^{-t} - 3e^{-3t})$

45 $e^{-t} - e^{-2t} + \frac{1}{2}(1 - 2e^{-(t-1)} + e^{-2(t-1)})u(t-1)$

47 $\frac{1}{2}t\sin t$

sec. 7.5
p 259

1 $\frac{1}{2}(5e^{-t} - 18e^{-2t} + 15e^{-3t})$

3 $\frac{1}{25}(3\cos t + 4\sin t - 3e^{-2t} - 10te^{-2t})$

5 $\frac{1}{20}[3e^t - 5e^{-t} + 2e^{-2t}(\cos t - 2\sin t)]$

7 $\dfrac{t}{4}(e^t - e^{-t})$

9 $y = \frac{1}{6}(2e^{2t} - 3e^t + e^{-t}) + \frac{1}{6}(3 + e^{2(t-2)} - 3e^{t-2} - e^{-(t-2)})u(t-2)$

11 $y = \frac{1}{4}[(1 + 2t + t^2)e^{-t} - \cos t - \sin t]$

13
$$x = -\frac{t^2}{2} + \frac{10t}{3} - \frac{14}{3} + \frac{14+4t}{3}e^{-t} + \left\{-\frac{5}{3} + (t-1) - \frac{1}{6}(t-1)^2 + \left[\frac{5}{3} + \frac{2(t-1)}{3}\right]e^{-(t-1)}\right\}u(t-1)$$

$$y = -\frac{t^2}{2} + \frac{4t}{3} - 2 + \frac{6+2t}{3}e^{-t} + \left[\frac{2}{3} - \frac{t-1}{3} + \frac{(t-1)^2}{6} - \left(\frac{2}{3} + \frac{t-1}{3}\right)e^{-(t-1)}\right]u(t-1)$$

sec. 7.6
p 269

3 $\dfrac{1}{1+s^2}\coth\dfrac{\pi s}{2}$

5 $\dfrac{1 - (1 + as)e^{-as}}{s^2(1 - e^{-2as})}$

7 $-\frac{1}{2}[\phi_5(\tau,1,2) - \phi_5(t,1,2)] + \frac{3}{5}[\phi_5(\tau,2,2) - \phi_5(t,2,2)] + \frac{3}{10}[\phi_9(\tau,0,1,2) - \phi_9(t,0,1,2)] + \frac{1}{10}[\phi_7(\tau,0,1,2) - \phi_7(t,0,1,2)]$

9 $[(-1)^n\phi_6(\tau,1,1) + \phi_6(t,1,1)] - [(-1)^n\phi_6(\tau,2,1) + \phi_6(t,2,1)] - [(-1)^n\phi_{12}(\tau,2,1) + \phi_{12}(t,2,1)]$

11 $y = -e^{-t} + 2e^{-3t} + [(-1)^n\phi_6(\tau,1,2) + \phi_6(t,1,2)] - [(-1)^n\phi_6(\tau,3,2) + \phi_6(t,3,2)]$

13 Proceeding naturally (though somewhat incautiously) we obtain at once

$$\mathcal{L}\{y\} = \left(\frac{1}{s} - \frac{s}{1+s^2}\right)\frac{1}{1+e^{-\pi s}}$$

and $y = \phi_2(t,\pi) - [(-1)^n\phi_8(\tau,0,1,\pi) + \phi_8(t,0,1,\pi)]$

However,

$$\phi_8(x,0,1,\pi) \sim \frac{\cos(x+\pi) + \cos x}{2(1+\cos\pi)} \sim \frac{0}{0} \qquad (1)$$

The reason for this is that the term of lowest frequency in the Fourier expansion of the driving function $f(t)$ duplicates a term in the complementary function $A \cos t + B \sin t$, and it, though not the rest of the terms, must be handled in a special way when the necessary particular integrals are obtained To circumvent this difficulty, it is convenient to write the second term in $\mathcal{L}\{y\}$ in the form

$$-\frac{s}{1+s^2}\left(1 - e^{-\pi s} + e^{-2\pi s} - \qquad\right)$$

Then, taking inverses, term by term, we find

$$y = \phi_2(t,\pi) - \phi_1(t,\pi)\cos t$$

sec. 7.7
p 280

1 $\dfrac{\sin 2t - 2t \cos 2t}{16}$ 3 $\delta(t) - 2e^{-2t}$

5 $\frac{1}{6}(\sin 3t + 3t \cos 3t)e^{-2t}$

7 $Y = \int_0^t \lambda e^{-a\lambda} f(t - \lambda)\, d\lambda = \int_0^t (t - \lambda)e^{-a(t-\lambda)} f(\lambda)\, d\lambda$

11 $A(t) = \dfrac{1 - 2e^{-t} + e^{-2t}}{2}$ $h(t) = e^{-t} - e^{\ 2t}$

13 $A(t) = \frac{3}{11} \times 10^{-3}(2e^{-2\,000t} - e^{\ 1\,000t/6})$

$h(t) = -\frac{12}{11}e^{-2\,000t} + \frac{1}{22}e^{-1{,}000t/6}$

$i_2(t) = \int_0^t (-\frac{12}{11}e^{-2\,000\lambda} + \frac{1}{22}e^{-1{,}000\lambda/6})E(t - \lambda)\, d\lambda$

19 (a) $\delta(t - a) * f(t) = \begin{cases} 0 & t < a \\ f(t - a) & 0 \leqq a \leqq t \end{cases}$

(b) $u(t - a) * f(t) = \begin{cases} 0 & t < a \\ \int_a^t f(t - \lambda)\, d\lambda & 0 \leqq a \leqq t \end{cases}$

(c) $t^m * t^n = \dfrac{m!n!}{(m + n + 1)!} t^{m+n+1}$

Chapter 8

sec. 8.2
p. 293

7 (b) No

sec. 8.3
p 299

1 $y(x,t) = \frac{1}{2}(1 - |x - at|)[u(x - at + 1) - u(x - at - 1)]$
$+ \frac{1}{2}(1 - |x + at|)[u(x + at + 1) - u(x + at - 1)]$

3 $y(x,t) = \frac{1}{2} \cos (x - at)[u(x - at + \pi/2) - u(x - at - \pi/2)]$
$+ \frac{1}{2} \cos (x + at)[u(x + at + \pi/2) - u(x + at - \pi/2)]$

5 $y(x,t) = \frac{1}{2}(x - at)e^{-(x - at)}u(x - at) + (x - at)e^{x-at}u(-x + at)$
$+ \frac{1}{2}(x + at)e^{-(x+at)}u(x + at) + (x + at)e^{x+at}u(-x - at)$

7 $\theta(x) = -(1 + \cos x) \quad x < 0$

9 $\phi(x) = (1/a) \cos x[u(x + \pi) - u(x - \pi)]$

11 If $f'' \equiv 0$, that is, if f is a linear function, then the given equation is identically satisfied without restriction on λ If f is an arbitrary, twice-differentiable, nonlinear function, substitution into the given equation yields $(a\lambda^2 + b\lambda + c)f'' = 0$, and this will be satisfied identically if and only if λ is a root of the quadratic equation $a\lambda^2 + b\lambda + c = 0$ According as $b^2 - 4ac$ is greater than, equal to, or less than zero, this equation will have two, one, or no real roots, as asserted

13 $x - at = c_1,\ x + at = c_2$

15 The two-dimensional wave equation has solutions of the form

$$u(x,y,t) = f(x - at) + F(x + at) + g(y - at) + G(y + at)$$

and also of the form

$$u(x,y,t) = f(x + y - \sqrt{2}\,at) + F(x - y - \sqrt{2}\,at) + g(-x + y - \sqrt{2}\,at) + G(-x - y - \sqrt{2}\,at)$$

where f, F, g, and G are arbitrary, twice-differentiable functions

sec. 8.4
p 309

1 $\int_0^l f(x)\,dx = 0$, $\int_0^l g(x)\,dx = 0$ Physically speaking, the first condition implies that the integrated initial angular displacement is zero, which will always be the case if the origin of θ is suitably chosen Since the shaft is of uniform cross section, the second condition implies that

$$\int_0^l I\dot{\theta}(x,0)\,dx = 0$$

which is precisely the statement that the total angular momentum of the shaft is initially (and hence permanently) zero In other words, $\int_0^l g(x)\,dx = 0$ implies that the vibration being studied is not superposed on a uniform rotation

3 (a) Yes, (b) yes, (c) yes, (d) yes, (e) no, (f) yes

5 $$\theta(x,t) = \sum_{n=0}^{\infty} A_n \sin\frac{n\pi x}{l}\cos\frac{n\pi at}{l} \qquad \text{where} \qquad A_n = \begin{cases} 0 & n \text{ even} \\ \dfrac{8l^2}{n^3\pi^3} & n \text{ odd} \end{cases}$$

7 Doubling the tension multiplies the frequency by $\sqrt{2}$ Because it is easier to change the length quickly and accurately than it is to change the tension

9 $$y(x,t) = \sum_{n=1}^{\infty} B_n \sin\frac{n\pi x}{l}\sin\frac{n\pi at}{l}$$

$$\text{where} \qquad B_n = \begin{cases} 0 & n \text{ even} \\ \dfrac{(-1)^{n+1}4l^2}{n^3\pi^3 a} & n \text{ odd} \end{cases}$$

11 $$y(x,t) = \sum_{n=1}^{\infty} \sin\frac{n\pi x}{l}\left(A_n\cos\frac{n\pi at}{l} + B_n\sin\frac{n\pi at}{l}\right)$$

$$\text{where} \qquad \begin{cases} A_1 = 1 & \\ A_n = 0 & n \neq 1 \end{cases} \qquad B_n = \frac{4l}{n^2\pi^2 a}\sin\frac{n\pi}{2}\sin\frac{n\pi}{4}$$

13 $$u(x,t) = \frac{u_0}{2} + \sum_{n=1}^{\infty} A_n e^{-n^2\pi^2 t/a^2 l^2}\cos\frac{n\pi x}{l}$$

$$\text{where} \qquad A_n = \begin{cases} 0 & n \text{ even} \\ -4u_0/n^2\pi^2 & n \text{ odd} \end{cases}$$

15 The normal modes of a uniform shaft of length l vibrating torsionally with its left end fixed and its right end free are given by the formula

$$\sin\frac{(2n - 1)\pi x}{2l} \qquad n = 1, 2, 3,$$

and the corresponding natural frequencies are $4l/(2n - 1)a$ For a similar shaft of length $2l$ vibrating torsionally with both ends fixed, the normal modes and natural frequencies are, respectively,

$$\sin\frac{m\pi x}{2l} \qquad \text{and} \qquad \frac{4l}{ma} \qquad m = 1, 2, 3, \ldots$$

Obviously, for every value of n, the nth natural frequency of the first shaft is the same as the natural frequency of order $m = 2n - 1$ of the second shaft Moreover the nth normal mode of the first shaft is clearly congruent to the portion of the $(2n - 1)$st normal mode of the second shaft which lies between 0 and l The converse is not true, however, for neither the normal modes nor the natural frequencies of even order of the shaft of length $2l$ correspond to possible free motions of the shaft of length l

sec. 8.5
p 327

3 $z_1 = 2.03,\ z_2 = 4.91,\ z_3 = 7.98,\ B_1 = 0.73,\ B_2 = -0.15,\ B_3 = 0.06$

5 $$u = \sum_{n=1}^{\infty} A_n e^{-z_n^2 t/a^2 l^2} \cos\left(\frac{z_n}{l} x\right) \qquad \text{where} \qquad A_n = \frac{200 \sin z_n}{z_n + \sin z_n \cos z_n}$$

and the z's are the roots of the equation $\cot z = \alpha z$

7 $$u = 70 + \sum_{n=1}^{\infty} A_n e^{-z_n^2 t/a^2 l^2} \cos\left(\frac{z_n}{l} x\right) \qquad \text{where} \qquad A_n = \frac{60 \sin z_n}{z_n + \sin z_n \cos z_n}$$

and the z's are the roots of the equation $\cot z = \alpha z$

11 In each case the natural frequencies are $\omega_n = z_n^2 a/l^2$, where z_n is the nth one of the roots of the indicated equation

(a) Hinged-hinged $\sin z = 0$, $X_n = \sin z_n x/l$

(b) Fixed-fixed $\cos z \cosh z = 1$

$$X_n = (\sin z_n - \sinh z_n)\left(\cos\frac{z_n x}{l} - \cosh\frac{z_n x}{l}\right) - (\cos z_n - \cosh z_n)\left(\sin\frac{z_n x}{l} - \sinh\frac{z_n x}{l}\right)$$

(c) Free-free. $\cos z \cosh z = 1$

$$X_n = (\sin z_n - \sinh z_n)\left(\cos\frac{z_n x}{l} + \cosh\frac{z_n x}{l}\right) - (\cos z_n - \cosh z_n)\left(\sin\frac{z_n x}{l} + \sinh\frac{z_n x}{l}\right)$$

(d) Fixed-hinged $\tan z = \tanh z$

$$X_n = (\sin z_n - \sinh z_n)\left(-\cos\frac{z_n x}{l} + \cosh\frac{z_n x}{l}\right) + (\cos z_n - \cosh z_n)\left(\sin\frac{z_n x}{l} - \sinh\frac{z_n x}{l}\right)$$

(e) Free-hinged $\tan z = \tanh z$

$$X_n = (\sin z_n + \sinh z_n)\left(\cos\frac{z_n x}{l} + \cosh\frac{z_n x}{l}\right) - (\cos z_n + \cosh z_n)\left(\sin\frac{z_n x}{l} + \sinh\frac{z_n x}{l}\right)$$

13 Assuming the beam to be hinged at $x = -l$ and at $x = l$ and to bear a mass M at $x = 0$, the frequency equation is

$$\sin z[2 \cos z \cosh z - kz(\cosh z \sin z - \sinh z \cos z)] = 0$$

where $z = \sqrt{\omega/a}\, l$ and k is the ratio of the attached mass to the mass of the beam. The factor $\sin z$ yields the frequencies for the modes of vibration in which the mass remains at rest and each half of the beam behaves as a simple hinged-hinged beam of length l

15 From Exercise 14, the normal modes are $X_n = \sin (z_n x/l)$, where the z's satisfy the equation $\cot z = rz$ and r is the ratio of the moment of inertia of the attached disk to the moment of inertia of the entire shaft From this it follows easily that $\int_0^l X_n X_m \, dx = r \sin z_n \sin z_m \neq 0$, since $\sin z_n \sin z_m = 0$ only if z_n or $z_m = k\pi$, and for these values $\cot z \neq rz$

17 Since $\int_{-\pi}^{\pi} \cos mx \cos nx \, dx = \begin{cases} 0 & m \neq n \\ \pi & m = n \end{cases}$, the system is orthogonal on the interval $(-\pi,\pi)$ However, since $\int_{-\pi}^{\pi} \sin x \cos nx \, dx = 0$ for all values of n, the system is not complete

sec. 8.6
p 336

1 Yes If the frequency of the impressed force is, say, $\omega_m = m\pi a/l$, then, for the term $C_m \sin (m\pi x/l) \sin (m\pi at/l)$ in $\phi(x) \sin (m\pi a/l)$, assume a term of the form $D_m [\sin (m\pi x/l)][t \cos (m\pi at/l)]$ in the series of particular integrals used in the second method

3 $Y = \left(\sum_{n=1}^{\infty} \frac{C_n}{\omega_n^2 - \omega^2} \sin \frac{n\pi x}{l} \right) \sin \omega t \qquad \text{where} \qquad C_n = \frac{4}{n\pi} \sin \frac{n\pi}{2} \sin \frac{n\pi}{4}$

5 If $n < c/2\pi a$, the time factor in the corresponding product solution is overdamped If $n > c/2\pi a$, the time factor in the corresponding product solution is underdamped If the string is acted upon by a forcing function $\phi(x) \sin \omega t$, the natural assumption $Y(x,t) = A(x) \sin \omega t + B(x) \cos \omega t$ leads to a pair of simultaneous differential equations for $A(x)$ and $B(x)$, and there is no simple extension of the concepts of magnification ratio and phase shift If the second method illustrated in Example 1 is employed and a particular integral of the form

$$\left(\sum_n D_n \sin \frac{n\pi x}{l} \right) \sin \omega t + \left(\sum_n E_n \sin \frac{n\pi x}{l} \right) \cos \omega t$$

is assumed, then for each value of n the corresponding term in the particular integral can be constructed using the concepts of magnification ratio and phase shift

7 The analysis proceeds very much as in Example 1 except that the normal modes of the problem are $\cos (n\pi x/l)$ rather than $\sin (n\pi x/l)$ Hence, $\phi(x)$ must be expressed as a half-range cosine expansion, and a similar assumption must be made for $\theta(x)$

9 $$y = \frac{-ag}{\rho A \xi^3} \left[\frac{\sin (z - \xi) + \sin \xi}{\sin z} - \frac{\sinh (z - \xi) + \sinh \xi}{\sinh z} - \xi(z - \xi) \right] \sin \omega t$$

where $\qquad z = \sqrt{\omega/a}\, l \qquad$ and $\qquad \xi = \sqrt{\omega/a}\, x$

11 $$u = \frac{1}{\pi} \int_{-\infty}^{\infty} \int_0^{\infty} e^{-\lambda^2 t/a^2} f(s) \cos \lambda s \cos \lambda x \, ds \, d\lambda$$

13 $$u = \frac{u_0}{\cosh 2\lambda - \cos 2\lambda} \left[\sin \omega t \cos \frac{\lambda x}{l} \cosh \left(2 - \frac{x}{l} \right) \lambda - \cos \omega t \sin \frac{\lambda x}{l} \sinh \left(2 - \frac{x}{l} \right) \lambda - \sin \omega t \cos \left(2 - \frac{x}{l} \right) \lambda \cosh \frac{\lambda x}{l} + \cos \omega t \sin \left(2 - \frac{x}{l} \right) \lambda \sinh \frac{\lambda x}{l} \right]$$

where $\lambda = al\sqrt{\omega/2}$. By applying this formula to each term in the Fourier expansion of an arbitrary periodic end condition, the steady-state temperature distribution produced by such an end condition can be determined

15 $$e(x,t) = \int_{-\infty}^{\infty} e^{-pt} \sin \lambda x[A(\lambda) \cos qt + B(\lambda) \sin qt]\, d\lambda + E_0e^{-ax} \cos (\omega t + bx)$$

where $A(\lambda) = \dfrac{1}{\pi} \displaystyle\int_0^\infty -E_0e^{-as} \cos bs \cos \lambda s\, ds$

and $B(\lambda) = \dfrac{1}{\pi q} \displaystyle\int_0^\infty \left[E_0\omega e^{-as} \sin bs + p \int_{-\infty}^{\infty} A(r) \sin rs\, dr\right] \sin \lambda s\, ds$

17 $$u = \sum_n \sum_m E_{mn}e^{-[(2m+1)^2+(2n+1)^2]\pi^2t/4a^2} \sin\left(\frac{2m+1}{2} \pi x\right) \sin\left(\frac{2n+1}{2} \pi y\right)$$

where $E_{mn} = 4 \displaystyle\int_0^1 \int_0^1 f(x,y) \sin\left(\frac{2m+1}{2} \pi x\right) \sin\left(\frac{2n+1}{2} \pi y\right) dx\, dy$

19 $$u = \sum_{n=1}^{\infty} A_n \sinh n\pi(1 - y) \sin n\pi x$$

where $A_n = \dfrac{2}{\sinh n\pi} \displaystyle\int_0^1 f(x) \sin n\pi x\, dx$

21 The problem can be solved by superposing the solutions to the problem defined by the following sets of boundary conditions

$u(x,0) = f_1(x) \qquad u(x,1) = u(0,y) = u(1,y) = 0$

$u(x,1) = f_2(x) \qquad u(x,0) = u(0,y) = u(1,y) = 0$

$u(0,y) = g_1(y) \qquad u(1,y) = u(x,0) = u(x,1) = 0$

$u(1,y) = g_2(y) \qquad u(0,y) = u(x,0) = u(x,1) = 0$

23 (a) $u = 100$ (as should be obvious)

(b) $u = 100x$ (as should be obvious)

(c) $$u = \sum_{n=1}^{\infty} \frac{400}{(2n-1)\pi} e^{-(2n-1)\pi y/2} \sin \frac{(2n-1)\pi x}{2}$$

25 $\omega_n = a\sqrt{m^2 + n^2}/2l$ ($m, n = 1, 2, 3,$), where l is the length of the edge of the membrane When the membrane is vibrating at a pure frequency the nodal curves are defined by the equation

$$A_{mn} \sin \frac{m\pi x}{l} \sin \frac{n\pi y}{l} + A_{nm} \sin \frac{n\pi x}{l} \sin \frac{m\pi y}{l} = 0$$

where A_{mn} and A_{nm} are arbitrary If either A_{mn} or A_{nm} is 0, the nodal curves are straight lines parallel to the edges of the membrane In general, however, the nodal curves are too complicated to describe explicitly

sec. 8.7
p. 343

3 (a) $$e(x,t) = \frac{ax}{2\sqrt{\pi}} \frac{e^{-a^2x^2/4t}}{t^{3/2}}$$

(b) $$e(x,t) = \int_0^t \left(1 - \operatorname{erf} \frac{ax}{2\sqrt{\lambda}}\right) E'(\lambda)\, d\lambda + \left(1 - \operatorname{erf} \frac{ax}{2\sqrt{t}}\right) E(0)$$

5 $$y = \frac{a^2}{(n^2\pi^2a^2/l^2) - \omega^2} \sin \frac{n\pi x}{l} \sin \omega t + \frac{al\omega}{n\pi(\omega^2 - n^2\pi^2a^2/l^2)} \sin \frac{n\pi x}{l} \sin \frac{n\pi at}{l}$$

The second term, whose frequency is different from that of the impressed force, is present only because friction has been neglected. Actually, it will die away rapidly in any realistic physical system The response of the string to a distributed force $f(x,t) = g(x) \sin \omega t$ or $F(x,t) = \sin (n\pi x/l)\, h(t)$, where $h(t)$ is periodic, can be found by first expanding $g(x)$ or $h(t)$, as the case may be, in a Fourier series and applying the result of the first part of the exercise to each term

7 $y = -\dfrac{gt^2}{2} + \dfrac{g}{2a^2}(at - x)^2 u(at - x)$

9 $\mathcal{L}\{\theta\} = \dfrac{a}{E_sJ}\,\dfrac{\sinh (sx/a)}{s \cosh (sl/a)}$ When $x = l$ this becomes $\dfrac{a}{E_sJ}\,\dfrac{1}{s}\tanh\dfrac{sl}{a}$, hence, in this case θ is the Morse dot function of period $4l/a$

Chapter 9

sec. 9.1
p 350

1 (a) $x = 0$, regular, (b) $x = 0$, irregular,
(c) $x = 0$, irregular, $x = 1$, regular
(d) $x = 1$, regular, $x = -1$, regular

3 $y = \sqrt{x}\left[1 - \dfrac{x}{(2!)^2} + \dfrac{x^2}{(3!)^2} - \dfrac{x^3}{(4!)^2} + \cdots\right] \quad x^2 < \infty$

A second series solution of this form cannot be found, since the indicial equation has equal roots

5 $y_1 = x^{-1}\left(1 - \dfrac{x^2}{2} + \dfrac{x^4}{2^2 2!\, 1 \cdot 5} - \dfrac{x^6}{2^3 3!\, 1 \cdot 5 \cdot 9} + \cdots\right) \quad x^2 < \infty$

$y_2 = \sqrt{x}\left(1 - \dfrac{x^2}{2 \cdot 7} + \dfrac{x^4}{2^2 2!\, 7 \cdot 11} - \dfrac{x^6}{2^3 3!\, 7 \cdot 11 \cdot 15} + \cdots\right) \quad x^2 < \infty$

7 The point at infinity is a regular singular point of the given differential equation

9 If the roots of the indicial equation differ by 1, the two roots lead to the same series unless $b_1(r - 1) + c_1 = 0$ Similar results hold if the roots differ by an integer greater than 1 For instance, if the roots differ by 2, the two roots lead to the same series unless

$$\begin{vmatrix} (r-1)(r-2) + b_0(r-1) + c_0 & b_1(r-2) + c_1 \\ b_1(r-1) + c_1 & b_2(r-2) + c_2 \end{vmatrix} = 0$$

sec. 9.2
p 356

3 If x_1 were a common zero of $J_\nu(x)$ and $J_{-\nu}(x)$, then, from the result of Exercise 2,

$$0 = -\frac{2}{\pi x_1}\sin \nu\pi$$

which is impossible, since ν is not an integer

sec. 9.4
p. 365

1 $y = \sqrt{x}\left[c_1 J_{1/(m+2)}\left(\dfrac{2}{m+2}x^{(m+2)/2}\right) + c_2 Y_{1/(m+2)}\left(\dfrac{2}{m+2}x^{(m+2)/2}\right)\right]$

3 $y = x^{-1}[c_1 J_0(2\sqrt{x}) + c_2 Y_0(2\sqrt{x})]$

5 $y = \sqrt{x}\, e^{-x}[c_1 J_{3/4}(\tfrac{1}{2}x^2) + c_2 J_{-3/4}(\tfrac{1}{2}x^2)]$

9 $y = c_1 J_0(2\sqrt{3x}) + c_2 Y_0(2\sqrt{3x}) + c_3 I_0(2\sqrt{3x}) + c_4 K_0(2\sqrt{3x})$

sec. 9.5
p 371

1 $J_5(x) = \left(\dfrac{384}{x^4} - \dfrac{72}{x^2} + 1\right)J_1(x) - \left(\dfrac{192}{x^3} - \dfrac{12}{x}\right)J_0(x)$

3 $-xJ_1(2x) + 2x^2 J_2(2x)$

19 (a) $\frac{1}{3}\{x^2[J_0(x)\sin x - J_1(x)\cos x] + xJ_1(x)\sin x\} + c$
(b) $\frac{1}{3}\{x^2[J_1(x)\cos x - J_0(x)\sin x] + 2xJ_1(x)\sin x\} + c$

23 $2\sqrt{x}\,J_1(\sqrt{x}) + c$

25 (a) $xI_1(x) + c$ (b) $x^2I_1(x) - xI_0(x) + \int I_0(x)\,dx + c$
(c) $xI_0(x) - \int I_0(x)\,dx + c$ (d) $x^2I_2(x) + c$

sec. 9.6
p 376

1 $1 = \sum_{n=1}^{\infty} \frac{2}{3\lambda_n J_1(3\lambda_n)} J_0(\lambda_n x)$ 3 $x = \sum_{n=1}^{\infty} \frac{-2}{\lambda_n J_0(2\lambda_n)} J_1(\lambda_n x)$

5 $x^2 = \sum_{n=1}^{\infty} \frac{-2}{\lambda_n J_1(\lambda_n)} J_2(\lambda_n x)$ 7 $1 = \sum_{n=1}^{\infty} \frac{-2}{3J_0(3\lambda_n)(1 + \lambda_n^2)} J_0(\lambda_n x)$

9 $J(\lambda)Y(5\lambda) - J(5\lambda)Y(\lambda) = 0$

sec. 9.7
p 386

1 $\frac{s}{(s^2 + a^2)^{3/2}}$ 3 $\frac{\lambda^n}{\sqrt{s^2 + \lambda^2}\,(\sqrt{s^2 + \lambda^2} + s)^n}$

5 (a) $1/\lambda$, (b) 0; (c) $1/\lambda^2$

7 $e^{-at} \int_0^t e^{at} J_0(bt)\,dt$

9 (a) $\frac{s}{(s^2 - \lambda^2)^{3/2}}$ (b) $\frac{\lambda}{(s^2 - \lambda^2)^{3/2}}$

11 $e^{t/2}I(t/2)$ 15 $\sin t - tJ_0(t)$

17 $-\frac{1}{\sqrt{t}} J_1(2\sqrt{t})$

21 $u = u_0 + (u_w - u_0)\frac{\cosh \sqrt{2h/wk}\,x}{\cosh \sqrt{2h/wk}\,a}$

where w is the constant thickness of the fin.

23 $u = u_0 + (u_e - u_0)\frac{K_1(\lambda\sqrt{R})I_0(\lambda\sqrt{x}) + I_1(\lambda\sqrt{R})K_0(\lambda\sqrt{x})}{K_1(\lambda\sqrt{R})I_0(\lambda\sqrt{r}) + I_1(\lambda\sqrt{R})K_0(\lambda\sqrt{r})}$

where $\lambda = 2\sqrt{2h/kw}$.

25 $\omega_1 = 2{,}400$ cycles/sec, $\omega_2 = 5{,}600$ cycles/sec

27 $x = (1 + \alpha t)\{c_1J_{1/3}[\frac{2}{3}\lambda(1 + \alpha t)^{3/2}] + c_2J_{-1/3}[\frac{2}{3}\lambda(1 + \alpha t)^{3/2}]\}$
where c_1 and c_2 are determined by the equations

$$c_1J_{1/3}(\tfrac{2}{3}\lambda) + c_2J_{-1/3}(\tfrac{2}{3}\lambda) = x_0$$

$$c_1J_{-1/3}(\tfrac{2}{3}\lambda) + c_2J_{1/3}(\tfrac{2}{3}\lambda) = 0$$

29 Using the first suggested method, the critical lengths are determined by the roots of the equation $J_{-1/3}(\frac{2}{3}al^{3/2}) = 0$ where $a = \sqrt{A\rho/EI}$ Using the second suggested method, the critical lengths are determined by the roots of the equation

$$\cos \tfrac{1}{2}\sqrt{3}\,\alpha l + \tfrac{1}{2}e^{-\alpha l/2} = 0 \qquad \text{where } \alpha = \sqrt[3]{\frac{A\rho}{EI}}$$

The first critical lengths in the respective cases are given by the formulas

$$l = 1.986\sqrt[3]{\frac{EI}{A\rho}} \qquad \text{and} \qquad l = 2.024\sqrt[3]{\frac{EI}{A\rho}}$$

31 $y = F_0 \tan\theta \left[x - \frac{\sqrt{x}\,J_0(2a\sqrt{x})}{aJ_1(2a\sqrt{l})}\right]$

33 $J_{3/2}\left(\frac{\omega R}{a}\right)J_{-3/2}\left(\frac{\omega r}{a}\right) - J_{3/2}\left(\frac{\omega r}{a}\right)J_{-3/2}\left(\frac{\omega R}{a}\right) = 0$

where $a^2 = \frac{E_s g}{\rho}\left(\frac{R - r}{l}\right)^2$

35 The natural frequencies are the values of ω determined by the equation

$$J_n\left(\frac{\omega b}{a}\right) = 0 \qquad n = 0, 1, 2, \ldots$$

where b is the radius of the drumhead.

37 $J_1(2\sqrt{\omega a l})I_2(2\sqrt{\omega a l}) - J_2(2\sqrt{\omega a l})I_1(2\sqrt{\omega a l}) = 0$

where $a^2 = 4\rho/Egk^2$

39 $u(r,\theta,z) = \sum_{n=0}^{\infty}\sum_{m=1}^{\infty}(A_{nm}\cos n\theta + B_{nm}\sin n\theta)\sinh(\lambda_{nm}z)J_n(\lambda_{nm}r)$

where $J_n(\lambda_{nm}b) = 0$ and

$$A_{nm} = \frac{2\int_0^b rG_n(r)J_n(\lambda_{nm}r)\,dr}{b^2\sinh(\lambda_{nm}h)J_{n+1}^2(\lambda_{nm}b)} \qquad G_n(r) = \frac{1}{\pi}\int_0^{2\pi} f(r,\theta)\cos n\theta\,d\theta$$

$$B_{nm} = \frac{2\int_0^b rH_n(r)J_n(\lambda_{nm}r)\,dr}{b^2\sinh(\lambda_{nm}h)J_{n+1}^2(\lambda_{nm}b)} \qquad H_n(r) = \frac{1}{\pi}\int_0^{2\pi} f(r,\theta)\sin n\theta\,d\theta$$

41 $u(r,t) = 100\frac{\ln r - \ln r_2}{\ln r_1 - \ln r_2}$

$$+ \sum_{n=1}^{\infty} e^{-\lambda_n^2 t/a^2}A_n[Y_0(\lambda_n r_1)J_0(\lambda_n r) - J_0(\lambda_n r_1)Y_0(\lambda_n r)]$$

where λ_n is the nth one of the roots of the equation

$Y_0(\lambda r_1)J_0(\lambda r_2) - J_0(\lambda r_1)Y_0(\lambda r_2) = 0$

and

$$A_n = \frac{-\int_{r_1}^{r_2} r[Y_0(\lambda_n r_1)J_0(\lambda_n r) - J_0(\lambda_n r_1)Y_0(\lambda_n r)]\,100\,\frac{\ln r - \ln r_2}{\ln r_1 - \ln r_2}\,dr}{\frac{r_2^2}{2}[Y_0(\lambda_n r_2)J_1(\lambda_n r_2) - J_0(\lambda_n r_2)Y_1(\lambda_n r_2)]^2 - \frac{r_1^2}{2}[Y_0(\lambda_n r_1)J_1(\lambda_n r_1) - J_0(\lambda_n r_1)Y_1(\lambda_n r_1)]^2}$$

43 $u(r,z) = \sum_{n=1}^{\infty} A_n J_0(\lambda_n r)\sinh(\lambda_n z)$, where λ_n is the nth one of the roots of the equation $cJ_0(\lambda b) - \lambda J_1(\lambda b) = 0$ and

$$A_n = \frac{2\int_0^b rf(r)J_0(\lambda_n r)\,dr}{b^2\sinh(\lambda_n h)[1 + (c/\lambda_n)^2]J_0^2(\lambda_n b)}$$

and c (instead of h) has been used for the parameter in the radiation law to avoid confusion with the height h of the cylinder.

45 $u(r,\theta) = \frac{200}{\pi}\theta + \frac{200}{\pi}\sum_{n=1}^{\infty}(-1)^n\left(\frac{r}{b}\right)^{2n}\frac{\sin 2n\theta}{n}$

sec. 9.8
p 398

1 (b) $x^2 = \dfrac{P_0(x) + 2P_2(x)}{3}$ $\quad x^3 = \dfrac{3P_1(x) + 2P_3(x)}{5}$

5 $P_n^{(k)}(0) = \dfrac{(2k)!}{2^k k!}\dbinom{-(2k+1)/2}{(n-k)/2}$ k, n both even or both odd

7 $u = \sum_{n=1}^{\infty} (Ar^n + B/r^{n+1})\, P_n(\cos\theta)$, where A and B are determined by the equations

$$Ab_1^n + \frac{B}{b_1^{n+1}} = \frac{n+1}{2}\int_0^\pi f_1(\theta)\sin\theta\, P_n(\cos\theta)\,d\theta$$

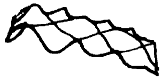
$$Ab_2^n + \frac{B}{b_2^{n+1}} = \frac{n+1}{2}\int_0^\pi f_2(\theta)\sin\theta\, P_n(\cos\theta)\,d\theta$$

9 $$H_{n1}(x) = a_1\left[1 - \frac{2n}{2!}x^2 + \frac{2^2 n(n-2)}{4!}x^4 - \qquad\right]$$

$$H_{n2}(x) = a_2\left[x - \frac{2(n-1)}{3!}x^3 + \frac{2^2(n-1)(n-3)}{5!}x^5 - \qquad\right]$$

The usual definitions are obtained by choosing a_1 and a_2 so that the coefficient of the highest power of x in each case is 2^n The orthogonality of the H's follows from the fact that the given differential equation can be written in the Sturm-Liouville form

$$\frac{d(e^{-x^2}y')}{dx} + 2ne^{-x^2}y = 0$$



Chapter 10

sec. 10.1
p 412

1 (a) 80, (b) 0, (c) 4

7 $D_n = aD_{n-1} - bcD_{n-2}$, if $a = 3$, $b = 2$, $c = 1$, then $D_n = 2^{n+1} - 1$

9 $\prod_{\substack{i,j=1\\ j>i}}^{n} (a_i - a_j)$ 15 Yes

sec. 10.2
p 428

1 The product of the given matrices is $\left\|\begin{matrix} 3 & 7 \\ 13 & 3 \end{matrix}\right\|$.

3 For $X = \left\|\begin{matrix} 1 & 2 \\ 2 & -1 \end{matrix}\right\|$ the given expressions each yield $\left\|\begin{matrix} 6 & -10 \\ -10 & 16 \end{matrix}\right\|$

For $X = \left\|\begin{matrix} 1 & 2 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{matrix}\right\|$ the given expressions each yield $\left\|\begin{matrix} 2 & -2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{matrix}\right\|$

The given relation is an identity for all square matrices X If X is not a square matrix the given expressions are meaningless.

7 AB_j, where B_j is the jth column vector of B

15 No; the individual submatrices must also be transposed, that is,

$A^T = \|a_{ji}{}^T\|$

17 The sum of the elements in any row, say the ith, is the sum of the probabilities that the system goes from the ith state to *some* state Since this is certain, the sum of these probabilities must be 1

19 $M = \frac{1}{8}\begin{Vmatrix} 4 & 3 & 1 \\ 3 & 4 & 1 \\ 3 & 3 & 2 \end{Vmatrix}$ $M^2 = \frac{1}{8^2}\begin{Vmatrix} 28 & 27 & 9 \\ 27 & 28 & 9 \\ 27 & 27 & 10 \end{Vmatrix}$

$M^3 = \frac{1}{8^3}\begin{Vmatrix} 220 & 219 & 73 \\ 219 & 220 & 73 \\ 219 & 219 & 74 \end{Vmatrix}$ $M^4 = \frac{1}{8^4}\begin{Vmatrix} 1{,}756 & 1{,}755 & 585 \\ 1{,}755 & 1{,}756 & 585 \\ 1{,}755 & 1{,}755 & 586 \end{Vmatrix}$

sec. 10.3
p. 436

3 (a) If A is nonsingular

5 The inverse of an upper (lower) triangular matrix is also upper (lower) triangular, and the elements on the principal diagonal of the inverse are the reciprocals of the corresponding elements in the original matrix.

7 The inverse of the coefficient matrix is $\frac{1}{8}\begin{Vmatrix} 1 & 3 & 1 \\ -3 & -1 & 5 \\ 2 & -2 & 2 \end{Vmatrix}$ Multiplying the matric form of the given equation by this matrix gives

$$X = \tfrac{1}{8}\begin{Vmatrix} 10 \\ 10 \\ 4 \end{Vmatrix}$$

9 Hint Verify that $(\text{adj } A)^{-1} = A/|A|$ Then, in the relation $\text{adj } B = |B|\, B^{-1}$, let $B = \text{adj } A$ and use the result of Exercise 8

13 $|K| = -k_{13}(k_1 + k_3)(k_{13} + k_{21}) - k_1k_3(k_{12} + k_{23}) - k_{12}k_{23}(k_1 + k_3)$, and if all the k's are positive, this is obviously a negative quantity.

15 K^{-1} = elasticity matrix $= -\frac{L^3}{162EI}\begin{Vmatrix} 2 & 5 & 8 \\ 5 & 16 & 28 \\ 8 & 28 & 54 \end{Vmatrix}$

K = stiffness matrix $= -\frac{162EI}{L^3}\begin{Vmatrix} 80 & -46 & 12 \\ -46 & 44 & -16 \\ 12 & -16 & 7 \end{Vmatrix}$

In this problem it is the elasticity matrix which is computed directly and the stiffness matrix which is computed as the inverse of the elasticity matrix In the example in the text it was the stiffness matrix which was computed directly and the elasticity matrix which was computed as the inverse of the stiffness matrix

sec. 10.4
p 443

1 No, because if all minors of order $\rho < r$ are zero, then, by expanding the minors of order r in terms of their ρth minors, we see that they too must vanish, contrary to hypothesis

3 (a) Rank $= \begin{cases} 3 & \lambda \neq \frac{1}{2}, 1, 2 \\ 2 & \lambda = 1, 2 \\ 1 & \lambda = \frac{1}{2} \end{cases}$

(b) Rank $= \begin{cases} 3 & \lambda \neq 1, 6 \\ 2 & \lambda = 1, 6 \end{cases}$

(c) Rank $= \begin{cases} 3 & \lambda \neq 1, \frac{19}{2} \\ 2 & \lambda = \frac{19}{2} \\ 1 & \lambda = 1 \end{cases}$

7 (a) T_1 column 2 − 2 column 1; T_2: column 3 − column 2;
T_3 column 1 + 2 column 2, T_4: column 1 − column 3;
T_5: column 2 − column 3, T_6: − column 2

(b) T_1: row 1 − row 3, T_2: row 3 − row 1, T_3 row 3 − row 2;
T_4: row 1 + row 3, T_5: row 2 + row 3, T_6 − row 3

$$P^{-1} = \begin{Vmatrix} 0 & -1 & 1 \\ -1 & 0 & 2 \\ 1 & 1 & -2 \end{Vmatrix}$$

9 A and B are equivalent since each is of rank 3 Hence in particular we can take $P = B$ and $Q = A^{-1}$.

sec. 10.5
p 459

5 If A is an (n,n) matrix of rank r, then $AX = O$ has $n - r$ linearly independent solution vectors, and so does $A^TX = O$. If A is an (m,n) matrix of rank r, then $AX = O$ has $n - r$ linearly independent solution vectors, but $A^TX = O$ has $m - r$ linearly independent solution vectors

7 $3X_1 + 4X_2 - 3X_3 + X_4 = O$, and from this each vector can immediately be expressed in terms of the other three

9 (a) $X = c_1\begin{Vmatrix} 5 \\ -6 \\ 2 \\ 1 \\ 0 \end{Vmatrix} + c_2\begin{Vmatrix} -4 \\ 3 \\ -1 \\ 0 \\ 1 \end{Vmatrix} + \begin{Vmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{Vmatrix}$ (b) $X = \begin{Vmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{Vmatrix}$

11 (a) $x_1 = -k,\ x_2 = k,\ x_3 = k$
(b) $x_1 = 17k,\ x_2 = -15k,\ x_3 = 13k,\ x_4 = 20k$

13 No, for instance, consider the matrix $\begin{Vmatrix} 1 & 1 \\ 1 & 2 \\ 0 & 0 \end{Vmatrix}$ Its rows are linearly dependent, but its columns are linearly independent

17 Any (p,q) matrix of rank r can be written as the product of a (p,r) matrix and an (r,q) matrix.

23 Note that $A^{p+1}X = O$ implies $A^{p+2}X = A^{p+3}X = \cdots = O$ Now consider the possibility that $c_1X + c_2AX + \cdots + c_pA^pX = O$ Multiplying on the left by A^p gives

$$c_1A^pX + c_2A^{p+1}X + \cdots + c_pA^{2p}X = O$$

which implies $c_1 = 0$ Similarly, it follows that $c_2 = c_3 = \cdots = c_p = 0$, which proves the given vectors are linearly independent

25 Hint. Use the result of Exercise 23

sec. 10.6
p 465

1 $X = c_1\begin{Vmatrix} -4 \\ 3 \end{Vmatrix} e^t + c_2\begin{Vmatrix} -5 \\ 3 \end{Vmatrix} e^{-2t} + \frac{1}{3}\begin{Vmatrix} 1 \\ 0 \end{Vmatrix} e^t$

3 $X = c_1e^{-2t}\left(\begin{Vmatrix} 0 \\ 1 \end{Vmatrix}\cos t + \begin{Vmatrix} -1 \\ 1 \end{Vmatrix}\sin t\right)$
$+ c_2e^{-2t}\left(\begin{Vmatrix} 1 \\ -1 \end{Vmatrix}\cos t + \begin{Vmatrix} 0 \\ 1 \end{Vmatrix}\sin t\right) + \begin{Vmatrix} -1 \\ 1 \end{Vmatrix} e^t$

5 $X = c_1\begin{Vmatrix} -2 \\ 1 \end{Vmatrix} + c_2\begin{Vmatrix} 1 \\ -3 \end{Vmatrix} e^{-5t} + \begin{Vmatrix} 1 \\ 0 \end{Vmatrix} + \begin{Vmatrix} -2 \\ 1 \end{Vmatrix} t$

7 $X = c_1e^{-t}\left(\begin{Vmatrix} 1 \\ -5 \end{Vmatrix}\cos 2t + \begin{Vmatrix} 3 \\ -3 \end{Vmatrix}\sin 2t\right)$
$+ c_2e^{-t}\left(\begin{Vmatrix} -3 \\ 3 \end{Vmatrix}\cos 2t + \begin{Vmatrix} 1 \\ -5 \end{Vmatrix}\sin 2t\right) + \frac{1}{6}\begin{Vmatrix} -1 \\ 5 \end{Vmatrix} e^{-t}$

Chapter 11

sec. 11.1
p 476

1 (a) Indefinite, (b) positive-definite, (c) negative-definite, (d) positive-semidefinite

5 (a) Minimum at $(0,-3)$, (b) critical point at $(-2,0,0)$, which is neither a maximum nor a minimum, (c) maximum at $(3,2,3)$, (d) minimum at $(1,0)$, $(-1,0)$ is a critical point which is neither a maximum nor a minimum, (e) minimum at $(1,1)$, $(0,0)$ is a critical point which is neither a maximum nor a minimum, (f) minima at $(\frac{1}{2}\pi + 2n\pi,\ \frac{1}{2}\pi + 2m\pi)$, maxima at $(\frac{1}{6}\pi + 2n\pi,\ \frac{1}{6}\pi + 2m\pi)$ and at $(\frac{5}{6}\pi + 2n\pi,\ \frac{5}{6}\pi + 2m\pi)$

9 $U_1 = \frac{1}{\sqrt{2}}\begin{Vmatrix} 0 \\ 0 \\ 1 \end{Vmatrix} \qquad U_2 = \frac{1}{\sqrt{2}}\begin{Vmatrix} 0 \\ 1 \\ 0 \end{Vmatrix} \qquad U_3 = \frac{1}{\sqrt{2}}\begin{Vmatrix} 2 \\ -1 \\ 0 \end{Vmatrix}$

sec. 11.2
p. 491

1 (a) $\lambda_1 = 1$, $\lambda_2 = 2$ (repeated, with a single independent characteristic vector); $X_1 = \begin{Vmatrix} 4 \\ 1 \\ -3 \end{Vmatrix}$, $X_2 = \begin{Vmatrix} 3 \\ 1 \\ -2 \end{Vmatrix}$

(b) $\lambda_1 = 1$ (repeated, with two independent characteristic vectors), $\lambda_2 = 2$, $(X_1)_1 = \begin{Vmatrix} 1 \\ 3 \\ 0 \end{Vmatrix}$, $(X_1)_2 = \begin{Vmatrix} 0 \\ -2 \\ 1 \end{Vmatrix}$, $X_2 = \begin{Vmatrix} 2 \\ 1 \\ 2 \end{Vmatrix}$

(c) $\lambda_1 = 0$, $\lambda_2 = 1$, $\lambda_3 = 2$, $X_1 = \begin{Vmatrix} 1 \\ 1 \\ 1 \end{Vmatrix}$, $X_2 = \begin{Vmatrix} 1 \\ -1 \\ 2 \end{Vmatrix}$, $X_3 = \begin{Vmatrix} 2 \\ 1 \\ 2 \end{Vmatrix}$

(d) $\lambda_1 = -1$, $\lambda_2 = 1$, $\lambda_3 = 4$, $X_1 = \begin{Vmatrix} 6 \\ 2 \\ -7 \end{Vmatrix}$, $X_2 = \begin{Vmatrix} 0 \\ 1 \\ -1 \end{Vmatrix}$, $X_3 = \begin{Vmatrix} 3 \\ 1 \\ -1 \end{Vmatrix}$

(e) $\lambda_1 = 1$ (repeated, with a single independent characteristic vector), $\lambda_2 = 6$, $X_1 = \begin{Vmatrix} 1 \\ 1 \\ -1 \end{Vmatrix}$, $X_2 = \begin{Vmatrix} 3 \\ 8 \\ 7 \end{Vmatrix}$

(f) $\lambda_1 = 1$ (repeated, with two independent characteristic vectors), $\lambda_2 = 6$, $(X_1)_1 = \begin{Vmatrix} 1 \\ 1 \\ 0 \end{Vmatrix}$, $(X_1)_2 = \begin{Vmatrix} 1 \\ 0 \\ 1 \end{Vmatrix}$, $X_2 = \begin{Vmatrix} 1 \\ 1 \\ 1 \end{Vmatrix}$

3 (a) $\lambda_1 = 1$ (repeated, with two independent characteristic vectors), $\lambda_2 = 2$, $(X_1)_1 = \frac{1}{2\sqrt{2}}\begin{Vmatrix} 1 \\ 0 \\ 1 \end{Vmatrix}$, $(X_1)_2 = \frac{1}{6\sqrt{2}}\begin{Vmatrix} -1 \\ 4 \\ 3 \end{Vmatrix}$, $X_2 = {}^1\!/_6\begin{Vmatrix} 1 \\ 2 \\ -3 \end{Vmatrix}$

(b) $\lambda_1 = 1$ (repeated, with two independent characteristic vectors), $\lambda_2 = 2$, $(X_1)_1 = {}^1\!/_3\begin{Vmatrix} 1 \\ 1 \\ 0 \end{Vmatrix}$, $(X_1)_2 = {}^1\!/_6\begin{Vmatrix} 1 \\ -2 \\ 3 \end{Vmatrix}$, $X_2 = {}^1\!/_6\begin{Vmatrix} -1 \\ 2 \\ 3 \end{Vmatrix}$

9 If the characteristic equation vanishes identically Example

$$A = \begin{Vmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & 1 & -1 \end{Vmatrix} \qquad B = \begin{Vmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{Vmatrix}$$

15 All characteristic vectors of the symmetric matrix $\begin{Vmatrix} 1 & i \\ i & -1 \end{Vmatrix}$ are proportional to $\begin{Vmatrix} 1 \\ -i \end{Vmatrix}$

sec. 11.3
p 503

1 (a) $P,Q = \begin{Vmatrix} 1 & 0 \\ -3 & 1 \end{Vmatrix}, \begin{Vmatrix} 1 & -2 \\ 0 & 1 \end{Vmatrix}, \begin{Vmatrix} 3 & -1 \\ -2 & 1 \end{Vmatrix}, \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix}$

(b) $P,Q = \begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix}, \begin{Vmatrix} 1 & 1 \\ 0 & 1 \end{Vmatrix}, \begin{Vmatrix} 1 & 0 \\ 3 & 1 \end{Vmatrix}, \begin{Vmatrix} 0 & 1 \\ 1 & 1 \end{Vmatrix}$

(c) $P,Q = \begin{Vmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ \frac{2}{3} & -\frac{1}{3} & 1 \end{Vmatrix}, \begin{Vmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}, \begin{Vmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ \frac{2}{3} & -\frac{1}{3} & 1 \end{Vmatrix}, \begin{Vmatrix} 0 & 1 & -1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}$

(d) $$P,Q = \begin{Vmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 3 & 1 \end{Vmatrix}, \begin{Vmatrix} 1 & 0 & -3 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{Vmatrix};$$

$$\begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 3 & 1 \end{Vmatrix}, \begin{Vmatrix} 1 & 0 & 3 \\ 1 & 1 & -2 \\ 0 & 0 & 1 \end{Vmatrix}$$

3 (a) $$\begin{Vmatrix} x_1 \\ x_2 \end{Vmatrix} = \begin{Vmatrix} 1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{6} \end{Vmatrix} \begin{Vmatrix} y_1 \\ y_2 \end{Vmatrix}$$

(b) $$\begin{Vmatrix} x_1 \\ x_2 \end{Vmatrix} = \begin{Vmatrix} 1/\sqrt{6} & 1/\sqrt{3} \\ -2/\sqrt{6} & 1/\sqrt{3} \end{Vmatrix} \begin{Vmatrix} y_1 \\ y_2 \end{Vmatrix}$$

(c) $$\begin{Vmatrix} x_1 \\ x_2 \end{Vmatrix} = \begin{Vmatrix} \frac{1}{2} & 3/\sqrt{12} \\ -\frac{1}{2} & 1/\sqrt{12} \end{Vmatrix} \begin{Vmatrix} y_1 \\ y_2 \end{Vmatrix}$$

(d) $$\begin{Vmatrix} x_1 \\ x_2 \end{Vmatrix} = \begin{Vmatrix} 1 & 1 \\ -1 & 0 \end{Vmatrix} \begin{Vmatrix} y_1 \\ y_2 \end{Vmatrix}$$

(e) $$\begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix} = \begin{Vmatrix} \frac{1}{2}\sqrt{6} & \frac{1}{4} & \frac{1}{4}\sqrt{3} \\ 1/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & -\frac{1}{2} & \frac{1}{2}\sqrt{3} \end{Vmatrix} \begin{Vmatrix} y_1 \\ y_2 \\ y_3 \end{Vmatrix}$$

(f) $$\begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix} = \begin{Vmatrix} \frac{1}{3} & -\frac{1}{3} & \frac{1}{6} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{Vmatrix} \begin{Vmatrix} y_1 \\ y_2 \\ y_3 \end{Vmatrix}$$

(g) $$\begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix} = \frac{1}{2} \begin{Vmatrix} 0 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{Vmatrix} \begin{Vmatrix} y_1 \\ y_2 \\ y_3 \end{Vmatrix}$$

(h) $$\begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix} = \begin{Vmatrix} \frac{1}{2}\sqrt{5} & -\frac{4}{3}\sqrt{21} & \frac{1}{3} \\ -\frac{1}{2}\sqrt{5} & -\frac{1}{3}\sqrt{21} & \frac{1}{3} \\ 0 & \frac{5}{3}\sqrt{21} & \frac{1}{3} \end{Vmatrix} \begin{Vmatrix} y_1 \\ y_2 \\ y_3 \end{Vmatrix}$$

5 (a) $$S = \begin{Vmatrix} 1 & 2 \\ 2 & 5 \end{Vmatrix} \qquad S^{-1} = \begin{Vmatrix} 5 & -2 \\ -2 & 1 \end{Vmatrix}$$

(b) $$S = \begin{Vmatrix} 1 & 1 \\ 1 & -1 \end{Vmatrix} \qquad S^{-1} = \frac{1}{2} \begin{Vmatrix} 1 & 1 \\ 1 & -1 \end{Vmatrix}$$

(c) $$S = \begin{Vmatrix} 1 & 1 \\ -2 & 1 \end{Vmatrix} \qquad S^{-1} = \frac{1}{3} \begin{Vmatrix} 1 & -1 \\ 2 & 1 \end{Vmatrix}$$

(d) $$S = \begin{Vmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \end{Vmatrix} \qquad S^{-1} = -\frac{1}{2} \begin{Vmatrix} 0 & -1 & -1 \\ -1 & 0 & 1 \\ -1 & 1 & 0 \end{Vmatrix}$$

(e) $$S = \begin{Vmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{Vmatrix} \qquad S^{-1} = \frac{1}{2} \begin{Vmatrix} 2 & 3 & -3 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{Vmatrix}$$

(f) $$S = \begin{Vmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 0 & -1 & 1 \end{Vmatrix} \qquad S^{-1} = \begin{Vmatrix} 0 & 1 & 1 \\ 1 & -1 & -2 \\ 1 & -1 & -1 \end{Vmatrix}$$

7 $$X = \frac{5}{3} \begin{Vmatrix} 1 \\ 1 \end{Vmatrix} \cos t + \frac{1}{2} \begin{Vmatrix} 2 \\ -1 \end{Vmatrix} \sin 2t - \frac{1}{3} \begin{Vmatrix} 2 \\ -1 \end{Vmatrix} \cos 2t$$

9 $$X = \frac{16}{45} \begin{Vmatrix} 3 \\ 3 \\ 2 \end{Vmatrix} \cos \frac{t}{2} - \frac{5}{45} \begin{Vmatrix} 3 \\ 0 \\ -10 \end{Vmatrix} \cos t + \frac{4}{45} \begin{Vmatrix} 3 \\ -12 \\ 2 \end{Vmatrix} \cos 2t$$

$$+ \frac{208}{180} \begin{Vmatrix} 3 \\ 3 \\ 2 \end{Vmatrix} \sin \frac{t}{2} + \frac{5}{180} \begin{Vmatrix} 3 \\ 0 \\ -10 \end{Vmatrix} \sin t + \frac{11}{360} \begin{Vmatrix} 3 \\ -12 \\ 2 \end{Vmatrix} \sin 2t$$

sec. 11.4
p. 515

5 Yes

9 For all values of a and b, the given equations are satisfied, respectively, by the following matrices

$$\text{(a)}\quad X = \begin{Vmatrix} a & -b \\ \dfrac{a^2-2a-3}{b} & 2-a \end{Vmatrix} \qquad \text{(b)}\quad X = \begin{Vmatrix} a & -b \\ \dfrac{a^2-4a+3}{b} & 4-a \end{Vmatrix}$$

$$\text{(c)}\quad X = \begin{Vmatrix} a & -b \\ \dfrac{a^2-4a-5}{b} & 4-a \end{Vmatrix} \qquad \text{(d)}\quad X = \frac{1}{a}\begin{Vmatrix} a & ab & 0 \\ 0 & 2a & 0 \\ -2 & 2b & 3a \end{Vmatrix}$$

15 (a) $X_1 = \begin{Vmatrix} 0 & 2 \\ -1 & 3 \end{Vmatrix}$ $\quad X_2 = \begin{Vmatrix} -1 & 4 \\ -2 & 5 \end{Vmatrix}$ $\quad X_3 = \begin{Vmatrix} 6 & -4 \\ 2 & 0 \end{Vmatrix}$ $\quad X_4 = \begin{Vmatrix} 5 & -2 \\ 1 & 2 \end{Vmatrix}$

(b) $X_1 = \begin{Vmatrix} -4 & 3 \\ -2 & 1 \end{Vmatrix}$ $\quad X_2 = \begin{Vmatrix} 4 & -9 \\ 6 & -11 \end{Vmatrix}$ $\quad X_3 = \begin{Vmatrix} -10 & 9 \\ -6 & 5 \end{Vmatrix}$ $\quad X_4 = \begin{Vmatrix} -2 & -3 \\ 2 & -7 \end{Vmatrix}$

(c) $X_1 = \begin{Vmatrix} 0 & 2 \\ -1 & 3 \end{Vmatrix}$ $\qquad$ (d) $X_1 = \begin{Vmatrix} 2 & -1 & -1 \\ -3 & 4 & 5 \\ 3 & -3 & -4 \end{Vmatrix}$

17 (a) $\begin{Vmatrix} 60 & 19 \\ -57 & -16 \end{Vmatrix}$ $\quad$ (b) $\begin{Vmatrix} -3 & -6 \\ 9 & 12 \end{Vmatrix}$ $\quad$ (c) $\begin{Vmatrix} 0 & -6 \\ 3 & 9 \end{Vmatrix}$

(d) $\begin{Vmatrix} -12 & -27 & -9 \\ 3 & 12 & 3 \\ 9 & 9 & 6 \end{Vmatrix}$ $\quad$ (e) $\begin{Vmatrix} 6 & 35 & 35 \\ 3 & 76 & 73 \\ -3 & -35 & -32 \end{Vmatrix}$

sec. 11.5
p 524

3 (a) $A^2 - A = O$, (b) $A^2 - I = O$, (c) $A^2 - 3A + 2I = O$, (d) $A^2 - 4A + 3I = O$

5 (a) $\dfrac{-A^2 + 8A}{35}$, (b) $\dfrac{-A^2 + 7A}{30}$, (c) $\dfrac{-A^2 + 3A}{10}$,

(d) $\dfrac{A^3 - 6A^2 + 65A}{90}$, (e) $\dfrac{A^3 - 10A^2 + 51A}{210}$

sec. 11.6
p 531

3 $e^{-A} = \begin{Vmatrix} 2e^{-1} - e^{-2} & 2e^{-1} - 2e^{-2} \\ -e^{-1} + e^{-2} & -e^{-1} + 2e^{-2} \end{Vmatrix}$

5 By the Cayley-Hamilton theorem,

$$e^A = I\left(1 + \frac{1}{2!} + \frac{1}{3!} + \frac{2}{4!} + \frac{3}{5!} + \frac{5}{6!} + \frac{8}{7!} + \frac{13}{8!} + \qquad\right)$$
$$+ A\left(1 + \frac{1}{2!} + \frac{2}{3!} + \frac{3}{4!} + \frac{5}{5!} + \frac{8}{6!} + \frac{13}{7!} + \frac{21}{8!} + \cdots\right)$$

By Sylvester's identity,

$$e^A = 2\sqrt{\frac{e}{5}}\sinh\frac{\sqrt{5}}{2}A + \sqrt{\frac{e}{5}}\left(\sqrt{5}\cosh\frac{\sqrt{5}}{2} - \sinh\frac{\sqrt{5}}{2}\right)I$$

9 The given matrices commute if and only if $y = 2x - 1$ When the matrices commute,

$\sin A = A \sin 1 \qquad \cos A = -A + I(1 + \cos 1)$

$\sin B = B \sin 1 \qquad \cos B = I \cos 1$

$$\sin(A+B) = (A+B)\frac{\sin 1 + \sin 2}{3} - I\frac{2\sin 1 - \sin 2}{3}$$

$$\sin(A-B) = (A-B)\sin 1$$

$$\cos(A+B) = I\cos 1$$

$$\cos(A-B) = (A-B)(\cos 1 - 1) + I$$

and, using these, the given identities can easily be verified

Chapter 12

sec. 12.1 p 543

1 (a) 3, 9, 13, −18, $10\mathbf{i} + 18\mathbf{j} + 16\mathbf{k}$, $-6/13$, $-6/9$, 131°49′, −90, $245\mathbf{i} + 210\mathbf{j} - 170\mathbf{k}$, 220, 0

(b) 7, 15, 11, 80, $72\mathbf{i} + 24\mathbf{j} - 12\mathbf{k}$, $-64/11$, $-64/15$, 40°22′, −636, $-210\mathbf{i} + 710\mathbf{j} + 425\mathbf{k}$, −1,272, 0

(c) 15, 15, 9, −40, $-75\mathbf{i} + 60\mathbf{j} + 30\mathbf{k}$, $62/9$, $62/15$, 96°7′, 915, $610\mathbf{i} + 100\mathbf{j} - 1{,}420\mathbf{k}$, 1,830, 0

7 Not necessarily, not necessarily, not necessarily, yes

9 $(17\mathbf{i} - 13\mathbf{j} + 8\mathbf{k})/\sqrt{522}$

11 No In fact, if $\mathbf{A} = 0$ and $\mathbf{B} = \mathbf{C} \neq 0$, then $\mathbf{A}\times\mathbf{B} = \mathbf{B}\times\mathbf{C} = \mathbf{C}\times\mathbf{A} = 0$, but $\mathbf{A} + \mathbf{B} + \mathbf{C} = 2\mathbf{C} \neq 0$

23 $$\mathbf{i} + 2\mathbf{j} + 3\mathbf{k} = \frac{-17\mathbf{A} + 14\mathbf{B} + 3\mathbf{C}}{33} = \frac{-11\mathbf{U} + 319\mathbf{V} + 143\mathbf{W}}{11}$$

where $\mathbf{U} = \dfrac{40\mathbf{i} + 45\mathbf{j} - 100\mathbf{k}}{330}$ $\mathbf{V} = \dfrac{2\mathbf{i} + 27\mathbf{j} + 28\mathbf{k}}{330}$ $\mathbf{W} = \dfrac{24\mathbf{i} - 6\mathbf{j} + 6\mathbf{k}}{330}$

25 No, because $\mathbf{F}\times\mathbf{R}$ is opposite to the direction in which $\mathbf{F}$ would cause a right-hand screw to advance

sec. 12.2 p. 549

3 (a) $\mathbf{U}\cdot\dfrac{d\mathbf{U}}{dt}\times\dfrac{d^3\mathbf{U}}{dt^3}$

(b) $\dfrac{d\mathbf{U}}{dt}\times\left(\dfrac{d\mathbf{U}}{dt}\times\dfrac{d^2\mathbf{U}}{dt^2}\right) + \mathbf{U}\times\left(\dfrac{d\mathbf{U}}{dt}\times\dfrac{d^3\mathbf{U}}{dt^3}\right)$

7 Yes, the tangents at $t = 0$ and $t = 1$ are parallel Yes, for all values of t the tangents at t and at $-t$ are parallel

9 $$\mathbf{P}(t) = \left(\frac{t^3}{3} + \frac{t^2}{2}\right)\mathbf{i} + \left(1 + \frac{t^5}{20}\right)\mathbf{j} + \left(\frac{t^3}{3} - \frac{t^4}{12} + 2\right)\mathbf{k}$$

15 $\mathbf{T} = \dfrac{\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}}{\sqrt{14}}$ $\mathbf{N} = \dfrac{-11\mathbf{i} - 8\mathbf{j} + 9\mathbf{k}}{\sqrt{266}}$ $\mathbf{B} = \dfrac{3\mathbf{i} - 3\mathbf{j} + \mathbf{k}}{\sqrt{19}}$

sec. 12.3 p. 558

1 3

3 (a) $yz + 3x^2 + 2xz - y^2$; (b) $-2yz\mathbf{i} + (xy - z^2)\mathbf{j} + (6xy - xz)\mathbf{k}$

5 $\pm\frac{1}{3}(2\mathbf{i} - 2\mathbf{j} - \mathbf{k})$ 15 $n = -3$

sec. 12.4 p 570

3 $\pi a^5/4$ 5 (a) 0; (b) $-8/15$

7 3 9 $\pi a^6/24$

11 The common value of the integrals is $\frac{2}{3}$

13 $(k/2)(x_1^2 + y_1^2)^{(n+1)/2} - (x_0^2 + y_0^2)^{(n+1)/2}$

17 Yes, provided the entire boundary is consistently traversed in the positive sense.

19 $\iint_R \left(\frac{\partial g}{\partial x}\frac{\partial f}{\partial y} - \frac{\partial g}{\partial y}\frac{\partial f}{\partial x}\right) dx\,dy$

sec. 12.5
p. 583

1 (a) $\frac{4}{3}$, (b) $\frac{11}{6}$; (c) $\frac{11}{6}$, (d) $\frac{11}{6}$

3 (a) 1, (b) 6, (c) $\pi/2$, (d) $\pi/3$

5
$$\iiint_V \left[u\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2}\right) - v\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right)\right] . V = \iint_S \left(u\frac{\partial v}{\partial n} - v\frac{\partial u}{\partial n}\right) dS$$

7
$$\int_C F_1\,dx + F_2\,dy + F_3\,dz = \iint_S \left[\left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}\right)\cos\alpha + \left(\frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}\right)\cos\beta + \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right)\cos\gamma\right] dS$$

9 The length of the curve No, because the vector **T** is defined only on the curve C

15 0 17 $\int_C \frac{1}{2} r^2\,d\mathbf{R}$

19 The common value of the integrals is $5\pi a^4/4$

21 The common value of the integrals is $\frac{3}{2}$

23 -2π

29 If O is a point at which the surface S has a tangent plane, the integral in Gauss' theorem is equal to 2π If O is a singular point on S, the integral may have any value between 0 and 4π

sec. 12.6
p 593

1 $-r^3/3$; $-\ln r$ 3 $\frac{2M}{a^2}(\sqrt{a^2 + z^2} - z)$

9
$$\mathbf{E} = \mathbf{i}\sum_{n=1}^{\infty}\left(\frac{4}{n\pi}\cos\frac{n\pi at}{l} + \frac{4l}{n^2\pi^2 a}\sin\frac{n\pi at}{l}\right)\sin\frac{n\pi y}{l} + \mathbf{k}\sum_{n=1}^{\infty}\left(\frac{4}{n\pi}\cos\frac{n\pi at}{l} - \frac{4l}{n^2\pi^2 a}\sin\frac{n\pi at}{l}\right)\sin\frac{n\pi y}{l} \qquad a^2 = \frac{1}{\mu\epsilon}$$

the summations extending only over the odd positive integers

Chapter 13

sec. 13.2
p 604

3 (a) $\hat{\mathbf{e}}_1 = \frac{\mathbf{i} - 2\mathbf{j} + \mathbf{k}}{3}$ $\hat{\mathbf{e}}_2 = \frac{-\mathbf{i} - \mathbf{j} + 2\mathbf{k}}{3}$ $\hat{\mathbf{e}}_3 = \frac{2\mathbf{i} + 5\mathbf{j} - 4\mathbf{k}}{3}$

$\hat{\mathbf{e}}^1 = 2\mathbf{i} + \mathbf{k}$ $\hat{\mathbf{e}}^2 = \mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$ $\hat{\mathbf{e}}^3 = \mathbf{i} + \mathbf{j} + \mathbf{k}$

(b) $\hat{\mathbf{e}}_1 = 2\mathbf{i} - \mathbf{j} - \mathbf{k}$ $\hat{\mathbf{e}}_2 = 3\mathbf{i} - 2\mathbf{j} - \mathbf{k}$ $\hat{\mathbf{e}}_3 = -7\mathbf{i} + 5\mathbf{j} + 3\mathbf{k}$

$\hat{\mathbf{e}}^1 = \mathbf{i} + 2\mathbf{j} - \mathbf{k}$ $\hat{\mathbf{e}}^2 = 2\mathbf{i} + \mathbf{j} + 3\mathbf{k}$ $\hat{\mathbf{e}}^3 = \mathbf{i} + \mathbf{j} + \mathbf{k}$

5 (b) $\cos\theta = \frac{\bar{U}^T\bar{G}^{-1}\bar{V}}{\sqrt{\bar{U}^T\bar{G}^{-1}\bar{U}}\sqrt{\bar{V}^T\bar{G}^{-1}\bar{V}}}$

sec. 13.3
p 618

1 (a) $f(x_1)\,\Delta x_1 + f(x_2)\,\Delta x_2 + f(x_3)\,\Delta x_3$

(b) $$\begin{aligned} &a_{11}x_1x_1 + a_{12}x_1x_2 + a_{13}x_1x_3 \\ &+ a_{21}x_2x_1 + a_{22}x_2x_2 + a_{23}x_2x_3 \\ &+ a_{31}x_3x_1 + a_{32}x_3x_2 + a_{33}x_3x_3 \end{aligned}$$

(c) $$\begin{aligned} &a_{11}x_1y_1 + a_{12}x_1y_2 + a_{13}x_1y_3 \\ &+ a_{21}x_2y_1 + a_{22}x_2y_2 + a_{23}x_2y_3 \\ &+ a_{31}x_3y_1 + a_{32}x_3y_2 + a_{33}x_3y_3 \end{aligned}$$

(d) $a_{11}x_1x_1 + a_{22}x_2x_2 + a_{33}x_3x_3$

(e) $$\begin{aligned} (a_1x^1 + a_2x^2 + a_3x^3)^2 &= (a_1x^1)^2 + (a_2x^2)^2 + (a_3x^3)^2 \\ &\quad + 2a_1a_2x^1x^2 + 2a_1a_3x^1x^3 + 2a_2a_3x^2x^3 \end{aligned}$$

(f) $$\begin{aligned} &\frac{\partial x^1}{\partial y^1}\frac{\partial y^1}{\partial x^k} z_1 + \frac{\partial x^1}{\partial y^2}\frac{\partial y^2}{\partial x^k} z_1 + \frac{\partial x^1}{\partial y^3}\frac{\partial y^3}{\partial x^k} z_1 \\ &+ \frac{\partial x^2}{\partial y^1}\frac{\partial y^1}{\partial x^k} z_2 + \frac{\partial x^2}{\partial y^2}\frac{\partial y^2}{\partial x^k} z_2 + \frac{\partial x^2}{\partial y^3}\frac{\partial y^3}{\partial x^k} z_2 \\ &+ \frac{\partial x^3}{\partial y^1}\frac{\partial y^1}{\partial x^k} z_3 + \frac{\partial x^3}{\partial y^2}\frac{\partial y^2}{\partial x^k} z_3 + \frac{\partial x^3}{\partial y^3}\frac{\partial y^3}{\partial x^k} z_3 \end{aligned}$$

(g) $$\begin{aligned} &\frac{\partial x^i}{\partial y^1}\frac{\partial y^1}{\partial x^1} z^1 + \frac{\partial x^i}{\partial y^2}\frac{\partial y^2}{\partial x^1} z^1 + \frac{\partial x^i}{\partial y^3}\frac{\partial y^3}{\partial x^1} z^1 \\ &+ \frac{\partial x^i}{\partial y^1}\frac{\partial y^1}{\partial x^2} z^2 + \frac{\partial x^i}{\partial y^2}\frac{\partial y^2}{\partial x^2} z^2 + \frac{\partial x^i}{\partial y^3}\frac{\partial y^3}{\partial x^2} z^2 \\ &+ \frac{\partial x^i}{\partial y^1}\frac{\partial y^1}{\partial x^3} z^3 + \frac{\partial x^i}{\partial y^2}\frac{\partial y^2}{\partial x^3} z^3 + \frac{\partial x^i}{\partial y^3}\frac{\partial y^3}{\partial x^3} z^3 \end{aligned}$$

7 (a) At each of the given points the lengths of the base vectors are $\mathbf{e}_1 = 1$, $\mathbf{e}_2 = 2$, $\mathbf{e}_3 = 1$

(b) At each of the given points the lengths of the reciprocal base vectors are $\mathbf{e}^1 = 1$, $\mathbf{e}^2 = \frac{1}{2}$, $\mathbf{e}^3 = 1$

9 At (2,0,1), $\mathbf{V} = -\mathbf{e}_1 + \frac{1}{2}\sqrt{3}\,\mathbf{e}_2$ At $\left(2, \frac{\pi}{3}, 1\right)$, $\mathbf{V} = \mathbf{e}_1 + \frac{1}{2}\sqrt{3}\,\mathbf{e}_2$

The length of $\mathbf{V}$ is 2

sec. 13.4
p 624

5 (a) No (b) No

sec. 13.5
p 628

3 (a) For a contravariant vector $\mathbf{V} = \xi^i\mathbf{e}_i$, the divergence is

$$\frac{1}{r^2 \sin\theta}\left[\frac{\partial(r^2 \sin\theta\ \xi^1)}{\partial r} + \frac{\partial(r^2 \sin\theta\ \xi^2)}{\partial\theta} + \frac{\partial(r^2 \sin\theta\ \xi^3)}{\partial\phi}\right]$$

If we let V^1, V^2, V^3 be the components of $\mathbf{V}$ along *unit* vectors in the directions of $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$, respectively, so that $V^1 = \xi^1$, $V^2 = r\xi^2$, $V^3 = r\sin\theta\ \xi^3$, the divergence appears in the more usual form

$$\frac{1}{r^2}\frac{\partial(r^2V^1)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(\sin\theta\ V^2)}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial V^3}{\partial\phi}$$

(b) For a covariant vector $\mathbf{V} = \xi_i\mathbf{e}^i$, the divergence is

$$\frac{1}{r^2 \sin\theta}\left[\frac{\partial(r^2 \sin\theta\ \xi_1)}{\partial r} + \frac{\partial(\sin\theta\ \xi_2)}{\partial\theta} + \frac{\partial[(1/\sin\theta)\xi_3]}{\partial\phi}\right]$$

If we let V_1, V_2, V_3 be the components of $\mathbf{V}$ along *unit* vectors in the directions of $\mathbf{e}^1$, $\mathbf{e}^2$, and $\mathbf{e}^3$, respectively, so that

$$V_1 = \xi_1 \qquad V_2 = \frac{\xi_2}{r} \qquad V_3 = \frac{\xi_3}{r\sin\theta}$$

the divergence appears in the more usual form

$$\frac{1}{r^2}\frac{\partial(r^2V_1)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(\sin\theta\ V_2)}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial V_3}{\partial\phi}$$

Chapter 14

sec. 14.2
p 635

5 $-19 - 22i$

7 $7i$

9 $\frac{1}{5}(-2 + 2i)$

13 $1 + i$, no

15 $x = 1, y = 2, x = 4, y = \frac{1}{8}$

sec. 14.3
p 641

1 (a) Rotation through $-90°$, (b) rotation through 45°

3 $\cos\frac{\pi}{4} + i\sin\frac{\pi}{4} = \frac{\sqrt{2} + i\sqrt{2}}{2}$

$\cos\frac{3\pi}{4} + i\sin\frac{3\pi}{4} = \frac{-\sqrt{2} + i\sqrt{2}}{2}$

$\cos\frac{5\pi}{4} + i\sin\frac{5\pi}{4} = \frac{-\sqrt{2} - i\sqrt{2}}{2}$

$\cos\frac{7\pi}{4} + i\sin\frac{7\pi}{4} = \frac{\sqrt{2} - i\sqrt{2}}{2}$

5 $\cos\frac{\pi}{6} + i\sin\frac{\pi}{6} = \frac{\sqrt{3} + i}{2}$

$\cos\frac{5\pi}{6} + i\sin\frac{5\pi}{6} = \frac{-\sqrt{3} + i}{2}$

$\cos\frac{9\pi}{6} + i\sin\frac{9\pi}{6} = -i$

7 $2^{1/6}(\cos 15° + i\sin 15°) = 1.084 + 0.291i$

$2^{1/6}(\cos 135° + i\sin 135°) = -0.794 + 0.794i$

$2^{1/6}(\cos 255° + i\sin 255°) = -0.291 - 1.084i$

9 $2^{2/5}(\cos 180° + i\sin 180°) = -1.320$

$2^{2/5}(\cos 108° + i\sin 108°) = -0.408 + 1.255i$

$2^{2/5}(\cos 36° + i\sin 36°) = 1.068 + 0.776i$

$2^{2/5}(\cos 324° + i\sin 324°) = 1.068 - 0.776i$

$2^{2/5}(\cos 252° + i\sin 252°) = -0.408 - 1.255i$

13 At the point $\frac{m_1z_1 + m_2z_2 + m_3z_3}{m_1 + m_2 + m_3}$

sec. 14.4
p 644

1 No

3 If and only if z_1 and z_2 have the same argument (or arguments differing by an integral multiple of 2π)

5 If and only if $y = \pm x$

7 The y-axis, the point (1,0), there is no locus

9 The set of all points on and within the parabola $y^2 = 2x - 1$

sec. 14.5
p. 649

1 $-2 - 3i$

3 $-iz^2 + 2z - 1$

5 (a) Unbounded, open, simply connected

(b) Bounded, closed, multiply connected

(c) Unbounded, open, multiply connected

(d) Unbounded, closed, simply connected

(e) Unbounded, neither open nor closed, simply connected

(f) Bounded, closed, simply connected

7 Along the parabolic paths $y = mx^2$ the function approaches the respective limits $\frac{m}{1 + m^2}$; hence, $\lim_{z\to 0} \frac{x^2y}{x^4 + y^2}$ does not exist

sec. 14.6
p 656

1 At $z = -1, \pm i$
3 Only at the origin; only at the origin, nowhere
5 If and only if u and v are constant
7 The values all lie on the circle $x = \dfrac{1 - m^2}{1 + m^2}$, $y = -\dfrac{2m}{1 + m^2}$, i.e , the circle $x^2 + y^2 = 1$

sec. 14.7
p 663

7 $(1 + i)^{1-i} = e^{(\ln\sqrt{2}+\pi/4+2n\pi)+i(-\ln\sqrt{2}+\pi/4+2n\pi)}$, and the arguments of the different values differ only by multiples of 2π
9 1   13 Yes
15 $z = \dfrac{\pi}{2} + 2n\pi + i\cosh^{-1} 3$ 17 $z = \ln 2 + (2n + 1)\pi i$
19 Since the complex numbers are not ordered, the inequalities appearing in Rolle's theorem are meaningless for complex variables
23 2

sec. 14.8
p 674

1 Along each path the integral is equal to $6 + 26i/3$
3 Along each path the integral is equal to $\frac{1}{6}(-1 + 5i)$
5 On the path $|z| = 3$ the absolute value of the integral is dominated by $\frac{3}{4}e^6$. (This is a very crude estimate, since by the methods of Chap 16 the exact value of the integral can be shown to be sin 2) On the path $|z| = \frac{1}{2}$ the integral, by Cauchy's theorem, is 0
7 (a) 0, (b) $\frac{1}{2}(3 + 2i)\pi$, (c) $\frac{1}{2}(-3 + 2i)\pi$
9 (a) $-3i\pi/2$, (b) $3i\pi/2$; (c) 0

Chapter 15

sec. 15.1
p. 685

1 The interior of the circle of radius 1 and center i
3 The parabola $y^2 = -1 - 2x$ and its interior
9 Yes, for instance, for all values of x, the series of continuous functions

$$-\frac{x}{1 + x^2} + \left[\frac{x}{1 + x^2} - \frac{2x}{1 + (2x)^2}\right] + \left[\frac{2x}{1 + (2x)^2} - \frac{3x}{1 + (3x)^2}\right] + \cdots$$

converges to the continuous function 0, although in the neighborhood of the origin the series does not converge uniformly In fact, $|R_n(x)| = \left|\dfrac{nx}{1 + (nx)^2}\right| < \epsilon$ implies that n must satisfy a requirement of the form $n > f(\epsilon)/|x|$

sec. 15.2
p 691

1 (a) $f(z) = -1 + 2z - 2z^2 + 2z^3 - \cdots \qquad |z| < 1$

(b) $f(z) = \dfrac{z - 1}{2} - \dfrac{(z - 1)^2}{4} + \dfrac{(z - 1)^3}{8} - \dfrac{(z - 1)^4}{16} + \cdots \qquad |z - 1| < 2$

3 (a) $f(z) = \dfrac{z}{2} - \frac{3}{4}z^2 + \frac{7}{8}z^3 - \frac{15}{16}z^4 + \cdots \qquad |z| < 1$

(b) $$f(z) = \left(\frac{1}{2} - \frac{1}{3}\right) - \left(\frac{1}{2^3} - \frac{1}{3^2}\right)(z - 2) + \left(\frac{1}{2^5} - \frac{1}{3^3}\right)(z - 2)^2 - \left(\frac{1}{2^7} - \frac{1}{3^4}\right)(z - 2)^3 + \left(\frac{1}{2^9} - \frac{1}{3^5}\right)(z - 2)^4 - \cdot \qquad |z - 2| < 3$$

5 $\sqrt{2}$ 7 $2\sqrt{2}$

sec. 15.3
p. 697

1 (a) $f(z) = \frac{1}{2} + \frac{3}{4}z + \frac{7}{8}z^2 + \frac{15}{16}z^3 + \cdots$

(b) $f(z) = \cdots - \frac{1}{z^3} - \frac{1}{z^2} - \frac{1}{z} - \frac{1}{2} - \frac{z}{4} - \frac{z^2}{8} - \frac{z^3}{16} - \cdots$

(c) $f(z) = \cdots + \frac{15}{z^5} + \frac{7}{z^4} + \frac{3}{z^3} + \frac{1}{z^2}$

(d) $f(z) = -\frac{1}{z-1} - 1 - (z-1) - (z-1)^2 - \cdots$

(e) $f(z) = \cdots + \frac{1}{(z-1)^4} + \frac{1}{(z-1)^3} + \frac{1}{(z-1)^2}$

(f) $f(z) = \frac{1}{z-2} - 1 + (z-2) - (z-2)^2 + (z-2)^3 - \cdots$

(g) $f(z) = \cdots + \frac{1}{(z-2)^4} - \frac{1}{(z-2)^3} + \frac{1}{(z-2)^2}$

3 $f(z) = -\frac{1}{z-i} - 2i + 3(z-i) + 4i(z-i)^2 - 5(z-i)^3 - \cdots \qquad 0 < |z-i| < 1$

$f(z) = \cdots - \frac{3}{(z-i)^5} - \frac{2i}{(z-i)^4} + \frac{1}{(z-i)^3} \qquad |z-i| > 1$

5 (a) 0, (b) $2\pi i$; (c) 0, (d) 0, (e) $2\pi i$, (f) $i\pi/6$

7 The argument is invalid because there is no value of z for which both series converge, that is, there is no value of z for which the two series are simultaneously valid

9 (c) $a_n = \frac{1}{2\pi}\int_0^{2\pi} \cosh\,(2\cos\theta)\cos n\theta\,d\theta$

(d) $a_n = \frac{1}{2\pi}\int_0^{2\pi} \cos\,(2\cos\theta)\cos n\theta\,d\theta$

Chapter 16

sec. 16.1
p 703

1 (a) $\frac{1}{2}$, (b) $\frac{1}{2}$

3 At $z = -1 + 2i$ the residue is $\frac{1}{4}(2 + i)$, at $z = -1 - 2i$ the residue is $\frac{1}{4}(2 - i)$

5 -1 7 $\frac{3}{10}$

9 (a) $2\pi i$, (b) 0; (c) 0, (d) 0; (e) 0,
(f) $\frac{1}{5}(5 + 2\sqrt{5})i\pi$

sec. 16.2
p 709

1 $\frac{2\pi}{1-p^2}$ 3 π

5 $\frac{2\pi}{b^2}(a - \sqrt{a^2 - b^2})$ 7 $\frac{\pi}{\sqrt{2}\,a^3}$

9 $\pi/3$ 11 $\pi/4a^3$

13 $\pi \cos ma\, e^{-mb}/b$ 15 $\pi(1 + am)e^{-am}/4a^3$

17 $\frac{\pi}{a^2 - b^2}\left(\frac{e^{-bm}}{b} - \frac{e^{-am}}{a}\right)$ 19 $\pi e^{-m/\sqrt{2}} \sin \frac{m}{\sqrt{2}}$

21 $\frac{\pi}{\sin a\pi}\left\{(b^2 + c^2)^{(a-1)/2} \sin\left[(a-1)\tan^{-1}\frac{-c}{-b}\right]\right\}$

23 $\frac{\pi}{3 \sin a\pi}\left(1 + 2\cos\frac{2\pi a}{3}\right)$

29 $a_n = \dfrac{(-1)^n 2}{\sqrt{a^2 - b^2}} \left[\tan\left(\dfrac{1}{2}\csc^{-1}\dfrac{a}{b}\right)\right]^n$

As $n \to \infty$, a_n approaches zero more rapidly than the reciprocal of any fixed power of n This is, of course, implied by Theorem 3, Sec 6 3, since all derivatives of the given function are everywhere continuous

sec. 16.3
p. 716

1 $\frac{1}{2}(e^{-t} - e^{-3t})$
3 $\frac{1}{2}\sin 2t$
5 $1 - \cos t$
7 $t \sin (2t/4)$
9 $e^{-2t}(2 - t) - 2e^{-3t}$

11 $f(x,t) = \dfrac{T_0}{E_s J}\left\{x - \dfrac{8l}{\pi^2}\sum_{n=0}^{\infty}\dfrac{\sin [(2n + 1)\pi x/2l] \cos [(2n + 1)\pi at/2l]}{(2n + 1)^3}\right\}$

13 $f(t) = 1 + \dfrac{4a}{\pi}\sum_{n=1}^{\infty}\dfrac{(-1)^{n+1}}{n}\sin\dfrac{n\pi t}{2a}$

15 $f(t) = 1 - \sum_{j=1}^{\infty}\dfrac{2J_0(\lambda_j r)e^{-\lambda_j t}}{\lambda_j J_1(\lambda_j)}$

where λ_j is the jth root of the equation $J_0(\lambda) = 0$.

sec. 16.4
p 728

1 $D_1 = 0$, $D_2 = -9$, $D_3 = -81$; therefore, there is at least one root with nonnegative real part. (Actually the roots are -1 92, 0 86 ± 1 94i)
3 $D_1 = 2$, $D_2 = 10$, $D_3 = 0$, $D_4 = 0$, therefore, there is at least one root with nonnegative real part. (Actually the roots are $\pm i\sqrt{2}$, $-1 \pm 2i$)

Chapter 17

sec. 17.1
p. 732

3 The transformation $w = f(\bar{z})$ is equivalent to the transformation $w = f(z)$ followed by a reflection in the real axis
5 Angles are not preserved
7 The equations of the transformation are

$u = x^4 - 6x^2y^2 + y^4 \qquad v = 4x^3y - 4xy^3$

The image of $x = 1$ is

$v^4 - 224uv^2 - 256u^3 - 2{,}176v^2 - 1{,}792u^2 - 2{,}084u + 4{,}096 = 0$

9 The equations of the transformation are

$u = 1 - \dfrac{y}{x^2 + y^2} \qquad v = -\dfrac{x}{x^2 + y^2}$

sec. 17.2
p. 737

1 (a) π/a; (b) $\pi/\sqrt{2}$
3 (a) $z = \pm 1$, (b) $|z^2 - 1| = \frac{1}{3}$;
(c) $x^2 - 2xy - y^2 = 1$, (d) $x^2 - y^2 = 1$
5 The images of the perpendicular lines $x = 1$ and $y = 0$ intersect at an angle of 45°.

sec. 17.3
p. 747

1 -1
3 If $(d - a)^2 + 4bc = 0$, the transformation leaves a single point invariant At least one point must be left invariant by any bilinear transformation.

5 $w = \dfrac{iz - 2}{z + 2}$, $3(u^2 + v^2) + 8u + 2v + 3 = 0$

7 $w = \dfrac{az + b}{cz + d}$, where a, b, c, d are all real and $ad - bc < 0$

9 $w = \dfrac{z^3 - i}{z^3 + i}$ **11** $w = -\left(\dfrac{z^4 - 1}{z^4 + 1}\right)^2$

17 $T = \dfrac{100}{\pi} \tan^{-1} \dfrac{3x^2y - y^3}{x^3 - 3xy^2}$ **19** $T = \dfrac{100}{\pi} \tan^{-1} \dfrac{1 - x^2 - y^2}{2y}$

sec. 17.4
p 753

1 $w = -z^2$ **3** $w = \sqrt{z}$ **5** $w = \sqrt{z^2 - 1}$

7 In the half plane onto which the given region is mapped by the mapping function $w = \sqrt{z^2 - 1} + \cosh^{-1} z$, the temperature is

$$T = \frac{1}{\pi} \tan^{-1} \frac{2xy}{x^2 - y^2 - 1}$$

This cannot be explicitly transformed back into an expression for the temperature in the original region in the w-plane

9 $w = i\pi + \tfrac{1}{2} \ln \dfrac{z + 1}{z - 1}$

Appendix
p 764

1 $r_1 = 3.325$, $r_2, r_3 = 1.338 \pm 0.632i$
3 $r_1 = 3.732$, $r_2 = -2.618$, $r_3 = -0.382$, $r_4 = 0.268$
5 $r_1 = 1.107$, $r_2 = -0.838$, $r_3 = r_4 = 0.500$, $r_5 = -0.270$
7 $r_1 = 4.966$, $r_2 = 2.450$, $r_3, r_4 = 0.612 \pm 1.129i$, $r_5 = -0.640$

ndex

The letter *e* after a page number refers to an exercise, the letter *n* to a footnote

Abel's identity, 32
Abscissa of convergence, 226
Absolute convergence, of infinite series, 677
  of Laplace transforms, 229
  of series of matrices, 527
Absolute value, of complex number, 641
  of vector, 4[illegible] 532
Acceleration, normal, 548
  tangential, 548
  vector, 548
Acceleration smoothing, 135, 142*e*
Addition, of complex numbers, 633
  of determinants, 408
  of infinite series, 678
  of matrices, 418
  of vectors, 533
Adjoint matrix, 430
Admittance, 169
  indicial, 273
Admittance matrix, 417
Advancing differences, 82
Ampere's law, 589
Amplitude, of complex number, 637
  of forced vibrations, 155–157, 329
  of free vibrations, 306, 307
  of *n*th harmonic, 197
Amplitude envelope, 153, 162
Amplitude modulation, 162, 165*e*
Amplitude spectrum, 213, 214
Analytic functions, 653
  derivatives of 653
  line integrals of, 667, 670
  mapping by, 732
  poles of, 699
    principal part at, 699
  properties of, 654, 655
  residues of, 700
Analytic functions, residues of, practical computation of, 702
  series of, 683, 684
  singular points of, 653, 699
Anharmonic ratio, 741
Annulus, 646
Antidifference, 87
  use in summing series, 88
Argand diagram, 636
Argument, of complex number, 637
    principal, 661
  principle of, 722
Augmented matrix, 447
Auxiliary equation, 38

Beams, bending of, 56
  bending moment on, 56
  deflection curve of, 56
  end conditions for, 323, 327*e*
  free vibration of, 287, 323
  load per unit length on, 56, 288
  neutral axis of cross section of, 56
  neutral surface of, 56
  shear on, 56
Beats, 162
Ber and bei functions, 365
Bernoulli numbers, 102
Bernoulli's equation, 21*e*
Bessel functions, asymptotic formulas for, 357*e*.
  behavior of, at origin, 356, 359
  differentiation of, 365, 366
  equations solvable in terms of, 363
  expansions in series of, 375, 380, 385
  of first kind, 353
    modified, 358

Bessel functions, of first kind, modified, asymptotic formulas for, 363*e*
  generating function for, 369
  integral formulas for, 371
  integration of, 368, 372*e*
  Laplace transforms of, 377, 386*e*
  of order $\pm\frac{1}{2}$, 364
  orthogonality of, 374
  recurrence relations for, 367
  of second kind, 354
    modified, 358
  of third kind, 354
  zeros of, 356, 359
Bessel's equation, equations reducible to, 363
  modified, 357
  of order ν, 351
    with a parameter, 351
  series solution of, 351
Bessel's inequality, 328*e*
Bilinear form, 469
Bilinear transformation, 737
Binomial coefficients, 83*n*
Binomial expansion, 695
Binormal, 550*e*
Boundary point, 646
Boundary-value problems, summary of, 326

Cantilever beam, end conditions for, 323
  vibration of, 323
Casoratti's determinant, 119*n*
Cauchy-Goursat theorem, 668
Cauchy-Riemann equations, 652
Cauchy's equation, 61*n*
Cauchy's inequality, 544*e* , 644*e*
  for $|f^{(n)}(z_0)|$, 673
Cauchy's integral formula, 669
  extensions of, 672
Cauchy's theorem, 667
Cayley-Hamilton theorem, 517
Center of gravity, 549
Central differences, 82
Channel, flow out of, 752
Characteristic curves, 301*e*
Characteristic equation, of boundary-value problem, 326
  of difference equation, 120
  of differential equation, 38, 52
  of matric equation, 477, 487
  of system of differential equations, 75, 462
Characteristic functions, 326
  orthogonality of, 320, 322
Characteristic polynomial, 477
Characteristic root, regular, 481
Characteristic-value problem, for matrices, 477, 487
  for partial differential equations, 461
Characteristic values, 326, 477
Characteristic vectors, 477
  orthogonality of, 484, 489
Characteristics, method of, 295*n* , 301*e*
Christoffel symbols, of first kind, 629
  of second kind, 629
Circle of convergence, 688
  use of, with series of real terms, 690
Closed formula, 110
Closed-loop system, 727
Closed set, 646
Closure, 318
Coefficient of correlation, 143*e*
Cofactors, 401
Columns, buckling of, 57
  critical loads for, 58
Complementary functions, of difference equations, 119
    table of, 121
  of differential equations, 35
    table of, 40
  of systems of differential equations, 76, 462
Complete solution, of difference equations, 119
  of differential equation, 5, 42
  of systems of algebraic equations, 448
  of systems of differential equations, 70, 76
Completeness, 317
Complex impedance, 169
Complex inversion integral, 225, 711
Complex number, 633
  absolute value of, 637, 641
  amplitude of, 637
  argument of, 637
    principal, 661
  components of, 634
  conjugate of, 634
  exponential form of, 658
  imaginary part of, 634
  logarithm of, 661
  modulus of, 637
  negative of, 634
  polar form of, 637
  powers of, 639, 661
  real part of, 634
  roots of, 639
  trigonometric form of, 637
Complex numbers, addition of, 634
  division of, 634
    in polar form, 638
  equality of, 634
  geometrical representation of, 636
  inequalities for absolute values of, 641
  multiplication of, 634
    in polar form, 638
  subtraction of, 634

- Complex plane, 636
  - integration in, 663–665
- Complex variable, 644
  - analytic function of, 653
  - exponential function of, 656
  - function of, 644
    - continuity of, 648
    - geometrical representation of, 729
    - regular point of, 653
    - singular point of, 653, 699
  - hyperbolic functions of, 660
    - inverse, 662
  - logarithmic function of, 660
  - trigonometric functions of, 659
    - inverse, 662
- Conductance, 169
- Conformal mapping, 722
  - behavior, of angles under, 735
    - of infinitesimal areas under, 734
    - of infinitesimal lengths under, 734
  - critical points of, 733
- Congruence transformation, 443
- Conjugate complex numbers, 634
- Connected set, 646
- Conservative field, 586
- Continuity equation of, 293, 555
  - of function of complex variable, 648
  - of sum of infinite series, 682
- Contour integral, 664
- Contravariant tensor, 620
  - covariant derivative of, 631
- Contravariant vector, 603, 620
- Convergence, abscissa of, 226
  - absolute, of infinite series, 677
    - of Laplace transform integral, 229
    - of series of matrices, 527
  - circle of, 688
  - conditional, 677
  - of Fourier series, 194
  - of improper integrals, 228
  - of infinite series, 676
    - ratio test for, 677
  - of Laplace transforms, 229, 231
  - in the mean, 318
  - radius of, 689
  - of series of matrices, 526
  - uniform, 678
    - of Laplace transform integral, 231
    - Weierstrass *M* test for, 681
- Convergence factor, 224
- Convolution integral, 271
- Coordinates, cylindrical, 566
  - generalized, 605
  - normal, 503
  - oblique, 597
  - spherical, 389
- Corrector formula, 110
- Cosine transform, 236
- Covariant tensor, 620
  - covariant derivative of, 631
- Covariant vector, 603, 620
- Cramer's rule, 453
- Critical damping, 152
- Critical points of conformal transformations, 733
  - behavior of angles at, 734
- Cross product, 531
- Cross ratio, 741
  - invariance of, 741
- Crosscuts, 647
- Curl, 554, 556
  - formulas for, 556
  - in generalized coordinates, 627
- Curvature, 548
  - radius of, 548
- Curve fitting, by factorial polynomials, 86
  - by harmonic analysis, 206
  - by Lagrange's formula, 94
  - by least squares, 135, 142*e*
  - by Newton's divided-difference formula, 91
  - by orthogonal polynomials, 133
- Curve smoothing, 135
- Curves, sectionally smooth, 559
  - simple closed, 568*n*
- Cylindrical coordinates, 56

- *D*, 36
- D'Alembert's solution of wave equation, 295
- Damped oscillation, 153
- Damping, critical, 152
  - viscous, 147
- Damping ratio, 157
  - relation of, to logarithmic decrement, 155
- Decibels, 155*n*
- Definite integrals, for Bessel functions, 371
  - differentiation of, 274*n*
  - evaluation of, by gamma functions, 239
    - by residues, 704
  - improper (*see* Improper integrals)
- Deformation of contours, principle of, 668
- Degrees of freedom, 145
- ∇, 552
- ∇^2, 588
- Δ, 81
- δ, 82
- δ-function, 276
- de Moivre's theorem, 639
- Dependence, linear, 444
- Determinants, addition theorem for, 408
  - definition of, 403
  - diagonal dominance in, 455*n*
  - differentiation of, 414*e*.
  - double subscript notation for, 401
  - elements of, 401

- Determinants, elements of, cofactors of, 401
  - minors of, 401
  - expansion of, Laplace's, 406
    - by cofactors, 403
  - Gramian, 460*e*
  - Jacobian, 609
  - *m*th-order minors of, 401
    - algebraic complements of, 401
    - complementary, 401
  - multiplication of, 411
  - properties of, 407–411
- Difference equations, 118
  - characteristic equation of, 120
  - complementary functions of, 119
    - table of, 121
  - complete solution of, 119
  - homogeneous, 118
  - nonhomogeneous, 118
  - order of, 118
  - particular solutions of, 121
    - table of, 122
  - solution of, 118
  - use of, in least squares, 143*e*
    - in summing series, 122
- Difference operators, 81
- Difference table, 80
- Differences, advancing, 82
  - central, 82
  - divided, 80
- Differential, exact, condition for, 583
  - total, 552
- Differential equation, 1
  - Bernoulli's, 21*e*
  - Bessel's, 351
    - modified, 357
  - Cauchy's, 61*n*
  - Euler's, 61
  - having a given general solution, 6
  - of heat flow, 290
  - Hermite's, 399*e*
  - Laguerre's, 399*e*
  - Laplace's, 291
    - in cylindrical coordinates, 294*e* , 382
    - in generalized coordinates, 627
    - in spherical coordinates, 389
  - Legendre's, 391
    - associated, 390
  - linear (*see* Linear differential equations)
  - nonlinear, 2
  - order of, 2
  - ordinary, 2
  - partial, 2
  - Poisson's, 588
  - solution of, 1
    - complete, 5
    - general, 5
- Differential equation, solution of, singular, 5
  - of vibrating beam, 288
  - of vibrating membrane, 285
  - of vibrating shaft, 287
  - of vibrating string, 284
- Differential equations, Cauchy-Riemann, 652
  - of electrical circuits, 150
  - first-order, exact, 14
    - existence theorem for, 5
    - homogeneous, 11
    - linear, 19
    - separable, 8
  - of higher order, 52
  - Maxwell's, 591
  - of mechanical systems, 150
  - numerical solution of, 108
    - Adams-Bashforth method, 116*e*
    - Adams-Moulton method, 117*e*
    - Euler's method, 112
    - Kutta's third-order approximation, 112
    - Milne's method, 108
    - modified Euler method, 112
    - Runge-Kutta method, 114
    - Runge's method, 112
  - of second order, homogeneous, 30
    - nonhomogeneous, 30
  - series solution of, 347
  - simultaneous (*see* Simultaneous differential equations)
  - solvable in terms of Bessel functions, 363
  - of transmission line, 292
- Differentiation, of analytic function, 653
  - of Bessel functions, 365, 366, 372*e*
  - of definite integral, 274*n*
  - of determinants, 414*e*
  - of Fourier series, 195
  - of improper integrals, 229
  - of infinite series, 684, 685
  - of Laplace transforms, 251
  - of matrices, 428*e*
  - numerical, 99, 136
  - of vector functions, 545
- Dimension of set of vectors, 456
- Directional derivative, 551
- Dirichlet conditions, 185
- Dirichlet's theorem, 185
- Distributions, theory of, 278*n*.
- Divergence, 554
  - formulas for, 556
  - in generalized coordinates, 624
- Divergence theorem, 572–574
- Divided differences, 80, 90*e*.
- Domain, 646
- Dot product, 418, 534
- Doublet function, 277
- Duhamel's formulas, 274

Dummy index, 610
Dyad, 535*n*.

E, 82
Eigenfunction, 326
Eigenvalue, 326
Eigenvector, 478*n*
Einstein summation convention, 610
Elastance, 165
Elasticity matrix, 435
Electrical circuits, differential equations of, 150
  forced vibrations of, 167
  free vibrations of, 166, 174
  laws of, 148, 149
Electrostatic field, 585
Elementary functions of complex variables, 656
Energy method, 59
Equivalence transformation, 443
Equivalent equations, 449*n*
Equivalent matrices, 438
Error function, 343
Error signal, 727
Essential singularity, 699
Euler-Maclaurin summation formula, 101
Euler's equation, 1
Euler's formulas, for cos θ and sin θ, 659
  for Fourier coefficients, 182
Euler's theorem on homogeneous functions, 14*e*
Even function, 189
  Fourier expansion of, 190
  Fourier integral of, 217
Exact differential equation, 15
Exponential function, 189
Exponential order, 226
Exterior point, 646

Factorial polynomials, 84
  expansions in terms of, 85
Falling bodies, 27*e*.
Faltung integral, 271
Faraday's law, 589
Feedback loop, 726
Feedback signal, 727
Field, conservative, 586
  electrostatic, 585
  gravitational, 585
  magnetic, 585
Field intensity, 586
Finite differences, 79
First-order reaction, 26
Forced motion, 151
Forced vibrations, of electric circuits, 68
  of mass-spring systems, 156, 161, 497
    magnification ratio for, 157
    phase angle for, 158
Fourier integrals, 211
  approximation by, when upper limit is finite, 219
  for even functions, 217
  exponential form of, 215
  initial conditions fitted by, 332
  as limit of Fourier series, 211
  for odd functions, 217
  relation to Laplace transforms, 222
  transform-pair forms of, 215, 217
  trigonometric form of, 216
Fourier series, 181
  alternative forms of, 196
  amplitude spectrum of, 213, 214
  coefficients of, 184
    behavior of, for large *n*, 194
  complex form of, 197
  convergence of, 194
  differentiation of, 195
  half-range, 192
    of even functions, 190, 196*e*
    of odd functions, 191, 196*e*
  initial conditions fitted by, 306, 308, 309
  integration of, 195
  *n*th harmonic of, 197
  periodic excitations represented by, 201, 204
  plots of partial sums of, 187
Fourier transform pair, for even function, 217
  for odd function, 217
  unilateral, 222
Fourier transforms, 221*e* , 222*e*
Fourier's law of heat conduction, 27*e*
Free motion, 151
  critically damped, 152
  overdamped, 151
  underdamped, 153
Free vibrations, amplitude of, 306, 308, 328
  of beams, 287, 323
  of electric circuits, 166, 174
  of mass-spring systems, 151
  of shafts, 302
  of strings, 298, 379
Frequency, effect of friction on, 154
  natural, of cantilever beams, 324
    of *LC* circuits, 166, 174
    of mass-spring systems, 154, 499
    of shafts, 306, 308, 309
    of strings, 299, 310*e*.
Frequency equation, 326
  determinantal, 499
Frequency ratio, 157

Friction, coefficient of, 164*e*.
  Coulomb, 164*e*
  effect of, on frequency, 154
  viscous, 146
Frobenius, method of, 347
Function, δ, 276
  entire, 691
  error, 343
  even, 189
  exponential, 657
  filter, 248
  gamma, 238
  generalized factorial, 238
  harmonic, 654
  holomorphic, 653
  homogeneous, 12
  impulse, 276
  integral, 691
  logarithmic, of complex variable, 660
    principal value of, 661
  Morse dot, 266
  odd, 190
  periodic, 182*n*.
  potential, 586
  rms value of, 200
  sine integral, 218
  staircase, 262
  transfer, 273, 727
  unit doublet, 277
  unit step, 237
Functions, analytic (*see* Analytic functions)
  ber and bei, 365
  Bessel (*see* Bessel functions)
  characteristic, 326
  complementary (*see* Complementary functions)
  conjugate harmonic, 654, 656*e*
  elementary, of complex variables, 656
  of exponential order, 226
  Hankel, 354
  hyperbolic, of complex variables, 660
    inverse, 662
  ker and kei, 362
  Legendre, of second kind, 391
  orthogonal (*see* Orthogonal functions)
  orthonormal, 315
  regular, 653
    piecewise, 226
  singularity, 277, 280*e*
  translated and "cut off," 247
  trigonometric, of complex variables, 659
    inverse, 662
  vector (*see* Vector functions)
Fundamental metric tensor, 621

Gamma function, 238
  graph of, 239
Gauss' law, for electric fields, 589
  for magnetic fields, 589
Gauss' reduction, 449
Gauss' theorem, 577, 585*e*.
Generalized coordinates, 605
  curl in, 627
  differential of arc in, 606
  divergence in, 627
  Laplacian in, 626
  length of vector in, 607
  local base vectors in, 606
  local reciprocal base vectors in, 606
  parametric curves in, 605
  transformations of, 609
Generalized functions, 278
Generalized orthogonality, 470
Generating function, for Bessel functions, 370
  for Legendre polynomials, 394
Gradient, 551
  geometrical properties of, 551, 552
Graeffe's root-squaring process, 755
Gram determinant, 460*e*.
Gramian, 460*e*
Gravitational field, 585
Gravitational potential, 588
Green's lemma, 567–570
Green's theorem, 575
Gregory-Newton formula, backward, 95
  forward, 94
Gregory's formula of numerical integration, 104

Half-range Fourier series, 189
Hankel functions, 354
Harmonic analysis, 206
Harmonic functions, 654
  conjugate, 654, 656*e*
Harmonics, 197
  higher, resonance with, 202
  spherical, 389
  surface, 390
  zonal, 392
Heat equation, 290
  solution of, 311, 333, 382, 397
  uniqueness of solutions of, 593
Heat flow, in cooling fins, 381, 387*e*
  in cylinders, 382
  differential equation of, 290
    uniqueness of solutions of, 593, 594*e*
  laws of, 27*e*., 288, 312
  in spheres, 397
  in thin rods, 311, 331
  in thin sheets, 333, 744
Heaviside's expansion theorems, 255
Hermite polynomials, 399*e*.
Hermite's equation, 399*e*.

- Hermitian form, 469
- Holomorphic function, 653
- Homogeneous differential equations, first-order, 11
  - higher-order, 30, 52
  - simultaneous, 74, 461
- Homogeneous functions, 12
  - Euler's theorem on, 14*e*
- Hyperbolic functions of complex variables, 660, 662

- Impedance, electrical, 167
    - complex, 169
    - parallel combinations of, 169
    - series combinations of, 169
  - mechanical, 167
- Improper integrals, continuity of, 228
  - convergence of, 228
  - differentiation of, 229
  - integration of, 229
  - principal value of, 706*n*
- Impulse function, 276
- Inconsistent equations, 450
- Independence, linear, 444
- Indicial admittance, 273
- Indicial equation, 348
- Inequalities, Cauchy's, 544*e* , 644*e*.
    - for $|f^{(n)}(z_0)|$, 673
  - for complex numbers, 641
  - for line integrals, 665
- Infinite series (*see* Series)
- Inner product, of tensors, 622
  - of vectors, 418, 534
- Integral, complex inversion, 225, 711
  - contour, 664
  - convolution, 271
  - of $J_0(x)$, 369*n*., 371*e*.
  - running, 105
  - surface, 565
  - volume, 566
  - (*See also* Definite integrals, Fourier integrals, Improper integrals, Line integrals, Particular integrals)
- Integrating factor, 17, 20
- Integration, of Bessel functions, 368, 372*e*
  - of Fourier series, 195
  - of improper integrals, 229
  - of infinite series, 683
  - of Laplace transforms, 252
  - line, 560
    - in complex plane, 664
  - numerical, 104
    - of differential equations, 108, 116*e*., 117*e*
- Integrodifferential equation, 148
- Interior point, 646
- Interpolation formulas, Gregory-Newton, backward, 95
    - forward, 94
  - Lagrange's, 94
  - Laplace-Everett, 97
  - Newton-Gauss, backward, 97
    - forward, 96
  - Newton's divided difference, 91
  - Stirling's, 97
- Inversion, 738

- $J_0(x)$, integral of, 369*n*., 371*e*.
- Jacobian, 609
  - of conformal transformation, 733
- Jacobian determinant, 609
- Jacobian matrix, 609
- Jordan canonical form, 497

- Ker and Kei functions, 362
- Kernel of transform, 236
- Kirchhoff's first law, 149
- Kirchhoff's second law, 148

- $\mathcal{L}$, 227
- $\mathcal{L}^{-1}$, 227
- Lagrange's identity, 543
- Lagrange's interpolation formula, 94
- Lagrange's reduction, 470
- Laguerre polynomials, 399*e*
- Laguerre's equation, 399*e*
- Lambert's law, 25*e*
- Laplace-Everett interpolation formula, 97
- Laplace transform pair, 225
- Laplace transforms, of Bessel functions, 377, 386*e*
  - convergence of, absolute, 229
    - uniform, 231
  - of derivatives, 234
  - differentiation of, 251
  - of elementary functions, 237
  - Heaviside's theorems on, 255
  - of integrals, 234
  - integration of, 252
  - inversion integral for, 225, 711
  - limit theorems for, 242, 243, 254*e*.
  - of periodic functions, 260
    - tables for, 266, 267
  - products of, 270
  - of products containing e^{-at}, 245
  - relation to Fourier integrals, 227
  - of singularity functions, 278, 280*e*
  - solution, of differential equations by, 235
    - of partial differential equations by, 338
  - of translated functions, 248

Laplace's equation, 291
  in cylindrical coordinates, 294e., 382
  in generalized coordinates, 626
  invariance under conformal transformation, 735
  relation to analytic functions, 654, 669
  in spherical coordinates, 389
Laurent's expansion, 692
  uniqueness of, 695
Least squares, 126
  acceleration smoothing by, 136
  curve smoothing by, 135, 142e
  dangers in logarithmic transformations in, 139
  relation to orthogonal functions, 131
  use in, of difference equations, 143e.
    of orthogonal polynomials, 130
    of Taylor series, 137
  velocity smoothing by, 135, 142e.
Legendre functions, 391
Legendre polynomials, 388
  algebraic form of, 392
  generating function for, 394
  orthogonality of, 397, 399e
  Rodrigues' formula for, 392
  series of, 398
  trigonometric form of, 395
Legendre's equation, 391
  associated, 390
    algebraic form of, 391
Leibnitz' rule, 274n
Lerch's theorem, 244n.
Lever surface, 552
Limit of function of complex variable, 648
Limit point, 646
Line integrals, in complex plane, 664
  conditions for independence of path, 579–582
  inequalities for, 665
  real, 560
    geometrical interpretation of, 562
  of vector functions, 560
Linear combination, 445
Linear dependence, 444
Linear differential equations, 2
  complementary function of, 35
  complete solution of, 33
  with constant coefficients, 35
    auxiliary equation of, 38
    characteristic equation of, 38
    higher order, 30, 52
    homogeneous, 30, 36
    nonhomogeneous, 30, 42
    particular integrals of, by Laplace transforms, 272
  exponents of, 348
  finding second solution of, 33
  first-order, 19
Linear differential equations, indicial equation of, 348
  ordinary point of, 345
  particular integral of, 35
    by variation of parameters, 49
  series solution of, 347
  singular point of, 345
    irregular, 346
    regular, 346
Linear equations, systems of, 447
    augmented matrix of, 447
    coefficient matrix of, 447
    complete solution of, 448
    equivalent, 449n.
    Gauss reduction for, 449
    homogeneous, 447
    inconsistent, 450
    nonhomogeneous, 447
    trivial solution of, 454
Linear fractional transformation, 737
Linear independence, 444
Linear transformation, 425
  matrix of, 425
Liouville's theorem, 691
Load per unit length, 56, 288
  relation to shear, 57
Logarithmic decrement, 155
  relation to damping ratio, 155
Logarithmic function, 660
  principal value of, 661
Lumped parameters, 145

M test, 681
Magnetic field, 585
Magnification ratio, 157
Mapping, 729
  conformal, 732
  isogonal, 735
Matric differential equations, 461
Matric equations, 447
  characteristic equation of, 477, 487
  characteristic values of, 477, 487
    reality of, 482
  characteristic vectors of, 477, 487
    independence of, 484, 486, 490
    normalized, 491
    orthogonality of, 484, 489
  solution of, 513
Matrices, addition of, 418
  conformable, 419
  conformably partitioned, 424
  diagonalization of, 493, 495
  elementary transformations of, 437
  equal, 415
  equivalent to diagonal matrix, 493
  multiplication of, 420, 422

Matrices, series of, 525
  similar, 443
    characteristic polynomials of, 479
  similar to diagonal matrix, 495
  subtraction of, 418
  transformations of, 492
    congruence, 443
    equivalence, 443
    orthogonal, 443
    similarity, 443
    unitary, 443
  transpose of products of, 423
Matrix, 415
  adjoint of, 430
  admittance, 417
  associate of, 416
  augmented, 447
  characteristic equation of, 477
  characteristic polynomial of, 477
  characteristic values of, 477
    multiple, 481
    reality of, 481
    regular, 481
  characteristic vectors of, 477
    independence of, 480, 484
    orthogonality of, 484
  coefficient, 447
  column, 415
  conjugate of, 416
  derivative of, 428*e*
  diagonal, 415
  elasticity, 435
  Hermitian, 416
  imaginary, 416
  inverse of, 430
  Jacobian, 609*n*.
  lower triangular, 415
  minimum polynomial of, 521
  minors of, 416
    principal, 416
  modal, 487
  nonsingular, 429
  *n*th power of, 506
  null, 416
  orthogonal, 435
  rank of, 437
    column, 456
    determinant, 437
    row, 456
    Sylvester's law of nullity for, 457
  real, 416
  row, 415
  scalar, 428*e*
  singular, 429
  skew-Hermitian, 416
  skew-symmetric, 416
  square, 415
    determinant of, 415
Matrix, square, functions of, 505
    square roots of, 514
    trace of, 479
    (*See also* Square matrix)
  stiffness, 434
  symmetric, 416
  transpose of, 415
  unit, 416
  unitary, 435
  zero, 416
Maxima and minima of functions of several variables, 474
Maxwell's equations, 591
Mechanical impedance, 167
Milne's method, 108
Minimum polynomial, 521
Minors, 401, 416
Möbius transformation, 737
Modal matrix, 487
Modified Bessel functions, of first kind, 358
  of second kind, 358
Modulus, of complex number, 637
  of elasticity, 294
  of spring, 59
Moment, bending, 56
  of force about a point, 545*e*
  vector, 545*e*
Motion, critically damped, 152
  forced, 151
  free, 151
  overdamped, 151
  steady-state, 160
  transient, 159
  underdamped, 153
Multiply-connected set, 646

Natural frequency, 154
Neighborhood, 646
Nepers, 155
Neutral axis, 56
Neutral surface, 56
Newton-Gauss interpolation formula, backward, 97
  forward, 96
Newton's divided-difference formula, 91
  remainder term in, 92
Newton's law of cooling, 27*e*
Newton's second law of motion, 27*e* , 146
  in torsional form, 26*e* , 146
Normal acceleration, 548
Normal coordinates, 503
Normal equations, 129
Normal modes, 326, 501
Null function, 316
Numerical methods, of differentiation, 99, 136
  of harmonic analysis, 206

Numerical methods, of integration, 103, 107*e*
  of solving differential equations, 108
  of solving equations, 314, 324, 735
Nyquist stability criterion, 726

Oblique coordinates, 597
  metrical properties of space in, 598, 601
  reciprocal base vectors in, 599
  reference vectors in, 597
    length of, 599
Odd function, 190
  Fourier expansion of, 191
  Fourier integral for, 217
Ohm's law, 167
Open formula, 110
Open-loop system, 726
Open set, 646
Operational calculus (*see* Laplace transforms)
Operators, D, 36
  ∇, 552
  ∇^2, 291, 558
  Δ, 81
  δ, 82
  E, 82
  equivalent, 83
  $\mathcal{L}$, 227
  $\mathcal{L}^{-1}$, 227
Order, of difference equation, 188
  of differential equation, 2
Ordinary point, 345
Orthogonal functions, 315
  closure of, 318
  completeness of, 317
  expansions in series of, 316
  least-square approximation by, 328*e*
Orthogonal polynomials, 130
  use of, in least squares, 131
    in smoothing of data, 135, 142*e*
Orthogonal trajectories, 28*e* , 655
Orthogonal transformations, 443
Orthogonal vectors, 419, 470, 534
Orthogonality with respect to, symmetric matrix, 470
  weight function, 316
Orthonormal functions, 315
Orthonormal vectors, 458
Osculating plane, 548, 550*e*

Partial differential equations, 2, 283
  elliptic, 301*e*
  hyperbolic, 301*e*
  parabolic, 301*e*
  solution of, by D'Alembert method, 294
    by Laplace transforms, 338
    by separation of variables, 302
Particular integrals, 35
  by Laplace transforms, 272
  for simultaneous differential equations, 76, 465
  by undetermined coefficients, 43
    table of, 46
  by variation of parameters, 49
Period, 182*n*.
Periodic function, 182*n*
Phase angle, 158
Poisson's equation, 588
Poisson's formula, 675*e*
Polar, 470
Pole, 699
  order of, 699
  principal part of $f(z)$ at, 699
Polynomials, factorial, 84
  finding zeros of, 755
  interpolation, 91, 94
  orthogonal, 130
Potential, 586
  gravitational, 588
Predictor formula, 110
Principle of argument, 722
Probability integral, 343*n*
Products, of complex numbers, 634, 638
  cross, 534
  of determinants, 411
  dot, 418, 534
  inner, 418
  of Laplace transforms, 270
  of matrices, 420, 422
  scalar, 418, 534
  scalar triple, 538
  of series, 678
  vector, 534
  vector triple, 541

Quadratic form, 466
  indefinite, 467
  kinetic energy, 503
  matrix of, 467
  negative, 466
  negative-definite, 466
    conditions for, 468
  nonsingular, 467
  polar of point with respect to, 470
  positive, 466
  positive-definite, 466
    conditions for, 468
  potential energy, 476*e* , 503
  reduction of, to sum of squares, 471
  semidefinite, 466
  singular, 467
Quality factor, 171*e*.

Radius, of convergence, 347, 689
  of curvature, 548
Rank, 437
  column, 456
  determinant, 437
  row, 456
  of tensor, 620, 621
Ratio test, 676
Reactance, 169
Region, 646
  boundary point of, 646
  bounded, 646
  closed, 646
  exterior point of, 646
  interior point of, 646
  multiply connected, 582
  open, 646
  simply connected, 582
  unbounded, 646
Regular function, 653
  piecewise, 226
Regular point, 653
Residue theorem, 701
Residues, 700
  calculation of, 702
  evaluation of definite integrals by, 704 708
Resonance, 2
  with higher harmonics, 202
Rodrigues' formula, 392
Root-coefficient relations, 719, 757
Routh-Hurwitz stability criterion, 721
Running integral, 105

Scalar, 418, 532
Scalar product, 418, 534
Scalar triple product, 538
Schmidt orthogonalization process, 458, 470
Schwarz-Christoffel transformation, 748
Self-conjugate form of equation of circle, 636*e* , 738
Separable differential equation, 8
Separation of variables, in first-order differential equations, 8
  in partial differential equations, 302
Sequence of matrices, convergence of, 525
  divergent, 525
Series, addition of, 678
  of Bessel functions, 375, 380, 385
  convergence of (*see* Convergence)
  differentiation of, 684, 685
  Fourier (*see* Fourier series)
  integration of, 683
  Laurent's, 692
  of Legendre polynomials, 398
  Maclaurin's, 689
  of matrices, 528
Series, multiplication of, 678
  of orthogonal functions, 318
  partial sums of, 676
  power, 689
  rearrangement of terms of, 677
  region of convergence of. 676
  remainder after *n* terms of, 676
  sum of, 676
    continuity of, 682
  summation of finite, 88, 89*e*
  Taylor's, 687
    circle of convergence of, 688
    radius of convergence of, 689
Set of vectors, dimension of, 456
  orthonormal, 458
  Schmidt orthogonalization process for, 458, 470
Shaft vibrations, longitudinal, 293*e*.
  torsional, 302
Shear, 56
  relation of, to bending moment, 57
Similarity transformation, 443
Simple closed curve, 568*n*
Simply connected set, 646
Simpson's rule, 107*e*
Simultaneous algebraic equations, 447
  consistency of, 450
  Gauss' reduction for, 449
  homogeneous, 447
    condition for nontrivial solution of, 454
    solution vectors of, 418, 447
  nonhomogeneous, 447
    augmented matrix of, 447
    solution of, by Cramer's rule, 453
Simultaneous differential equations, 66, 461
  characteristic equation of, 75, 462
  complementary function for, 76, 463
  complete solution of, 76, 463
  in matrix form, 461
  particular integral of, 76, 465
  reduction to single equation, 67
Sine integral function, 218
Sine transform, 236
Singular point, of analytic function, 653
    essential, 699
    isolated, 699
  of differential equation, 346
    irregular, 346
    regular, 346
  residue at, 700
Singularity functions, 277
Specific heat, 189
Spherical coordinates, 389
  differential of arc length in, 615
Spherical harmonic, 389
Spring modulus, 59

Square matrix, 415
  characteristic equation of, 477
  characteristic values of, 477
  characteristic vectors of, 477
  integral powers of, 505
  minimum polynomial of, 521
  polynomial annihilator of, 521
  polynomial equations in, 513
  polynomial functions of, 506
    characteristic vectors of, 510
  power series in, 528
  rational functions of, 508
    characteristic values of, 510
  trace of, 479
  (*See also* Matrix, square)
Stability criteria, 716
  for cubic equations, 719
  Nyquist, 726
  Routh-Hurwitz, 721
Static deflection, 157
Steady state, 160
Stefan's law, 312
Stiffness matrix, 434
Stirling's interpolation formula, 97
Stoke's theorem, 577-579
Stream lines from channel, 752
String, forced vibrations of, 329
  traveling waves on, 296, 298, 339
  vibration of, 283, 379
Sturm-Liouville theorem, 320
  extension to fourth-order systems, 322
Submatrices, 416
Summation of finite series, 88, 89*e*
Surface harmonic, 390
Susceptance, 169
Systems, with one degree of freedom, 145
  with several degrees of freedom, 145, 171
    use of difference equations in, 123, 174

Tangential acceleration, 548
Taylor's series, 687
  use of, in least squares, 137
Taylor's theorem, 686
Telegraph equations, 292
Telephone equations, 292
Tensor, alternating, 621
  or arbitrary rank, 621
  components of, 620
  contraction of, 622
  contravariant, of rank 1, 620
    of rank 2, 620
  covariant, of rank 1, 620
    of rank 2, 620
  covariant derivative of, 631, 632
  fundamental metric, 621
  mixed, of rank 2, 621
  of rank zero, 620
Tensor, skew-symmetric, 621
  symmetric, 621
Tensors, equal, 621
  inner product of, 622
  outer product of, 621
  quotient law for, 623
  sum of, 621
Theorem, binomial, 695
  Cauchy-Goursat, 668
  Cauchy's, 667
  de Moivre's, 639
  divergence, 572
  Euler's, 14
  Gauss', 577, 585*e*
  Green's, 575
  Heaviside's, 255
  Lerch's, 244*n*
  Liouville's, 691
  maximum modulus, 674
  Morera's, 672
  Parseval's, 318
  residue, 701
  Stoke's, 577-579
  Sturm-Liouville, 320
    extended, 622
  Taylor's, 686
Theory of distributions, 278*n*
Thermal conductivity, 289
Torricelli's law, 23
Torsion of space curve, 550*e*
Torsional rigidity, 286
Total differential, 14, 552
Trajectories, orthogonal, 28*e* , 655
Transfer function, 273, 727
Transformations of matrices, 443
Transforms, cosine, 236*e*
  Fourier, 221*e*
  Laplace (*see* Laplace transforms)
  sine, 236*e*
  (*See also* Fourier transform pair)
Transient, 159
Transition probabilities, matrix of, 426
Transmission line, equations for, 291
  steady-state behavior of, 332
  transient behavior of, 342
Trapezoidal rule, 105
  use of, in running integration, 105
Trigonometric functions of complex variables, 659, 662
Trivial solution, 454

Umbral index, 610
Undetermined coefficients, method of, 43, 121
Uniform convergence, of infinite integrals, 228
  of infinite series, 678

Uniform convergence, of Laplace transform integral, 231
Unilateral Fourier transform pair, 222
Unit doublet, 277
Unit impulse, 276
Unit step function, 237
Unit triplet, 278
Unit vectors, 419, 533
Unitary transformations, 443

Variation of parameters, 49
Vector, 417, 532
  absolute value of, 418, 532
  components of, 418
  curl of, 554
  divergence of, 554
  length of, 418
    generalized, 470
  negative of, 533
  product of scalar and, 533
  unit, 419, 533
  zero, 533
Vector acceleration, 548
Vector angular velocity, 555
Vector functions, 545
  derivative of, 545
  differential of, 545
  line integral of, 560
  surface integral of, 565
Vector moment, 545*e*
Vector product, 534
Vector triple product, 541
Vector velocity, 548
Vectors, addition of, 533
  characteristic, of square matrix, 478
  contravariant representation of, 603
  covariant representation of, 603
  cross product of, 534
Vectors, difference of, 533
  dot product of, 418, 534
  equal, 533
  inner product of, 418, 534
  normalized, 470
  orthogonal, 419
  orthonormal, 458
  reciprocal sets of, 544*e*, 599
  in same direction, 419
  scalar product of, 534
  solution, 418, 447, 477
    orthogonality of, 484, 488, 500
Velocity smoothing, 134, 142*e*
Vibrations, amplitude modulation of, 162
  of beams, 287, 323
  damped, 153
  of electric circuits, 165, 174
  forced (*see* Forced vibrations)
  free (*see* Free vibrations)
  of membranes, 284
  normal modes of, 326, 501
  of shafts, 285, 302
  of strings, 282, 298, 329, 379
Volume integral, 566

Wave equation, one-dimensional, 284
    D'Alembert solution of, 295
  two-dimensional, 285
Weierstrass *M* test, 681
Work, 563
Wronskian, 31, 52*n*

Zeros, of Bessel functions, 353, 356, 357*e*., 359
  within given contour, 722
Zonal harmonic, 392